The Great Courses®

Topic
Philosophy
& Intellectual History

Subtopic
Understanding
the Mind

Mind-Body Philosophy

Course Guidebook

Professor Patrick Grim
Stony Brook University

PUBLISHED BY:

THE GREAT COURSES
Corporate Headquarters
4840 Westfields Boulevard, Suite 500
Chantilly, Virginia 20151-2299
Phone: 1-800-832-2412
Fax: 703-378-3819
www.thegreatcourses.com

Printed in the United States of America

Patrick Grim

B.Phil., Ph.D.

Distinguished Teaching Professor of Philosophy Emeritus

Stony Brook University

Patrick Grim is the Distinguished Teaching Professor of Philosophy Emeritus at Stony Brook University. Having graduated with highest honors in both Anthropology and Philosophy from the University of California, Santa Cruz, Professor Grim was named a Fulbright Fellow to the University of St. Andrews, Scotland, from which he received his B.Phil. He received his Ph.D. from Boston University with a dissertation on ethical relativism and spent a year as an Andrew W. Mellon Faculty Fellow at Washington University.

Professor Grim has received the president's and chancellor's awards for excellence in teaching at Stony Brook University. He was named Marshall M. Weinberg Distinguished Visiting Professor at the University of Michigan in 2006 and visiting fellow at the Center for Philosophy of Science at the University of Pittsburgh in 2007. He has also been a frequent visiting scholar and professor at the Center for the Study of Complex Systems at the University of Michigan.

Professor Grim has published extensively in computational modeling and on such topics as theoretical biology, linguistics, decision theory, and artificial intelligence. His work spans ethics; philosophical logic; game theory; contemporary metaphysics; and philosophy of science, law, mind, language, and religion.

Professor Grim is the author of *The Incomplete Universe: Totality, Knowledge, and Truth* and the coauthor of *The Philosophical Computer: Exploratory Essays in Philosophical Computer Modeling* (with Gary Mar and Paul St. Denis); *Beyond Sets: A Venture in Collection-Theoretic Revisionism* (with Nicholas Rescher); and *Reflexivity: From Paradox to Consciousness* (also with Nicholas Rescher). He is the editor of Mind and Consciousness: *5 Questions* and *Philosophy of Science and the Occult* as well as a founding coeditor of more than 30 volumes of *The Philosopher's Annual*, an anthology of the best articles published in philosophy each year.

Professor Grim has taught three other Great Courses: *The Philosopher's Toolkit: How to Be the Most Rational Person in Any Room*; *Philosophy of Mind: Brains, Consciousness, and Thinking Machines*; and *Questions of Value*. ■

Table of Contents

Introduction

Lecture Guides

References

Acknowledgments

Professor Grim is grateful to his undergraduates studying philosophy of mind for research input, and to his graduate students in his seminar on the brain and mind for extended discussion.

He is extremely grateful to L. Theresa Watkins, his collaborator in all things, for good ideas, for extensive editing of both form and content, and for loving encouragement throughout.

Mind-Body Philosophy

How can three and a half pounds of gray matter in our skulls produce the world of subjective experience? What is the relation between minds and bodies—between the mental and the physical? How does the brain produce the phenomena of consciousness in memory, emotion, perception, altered states, and our sense of ourselves? Questions of bodies and minds have been topics of intense concentration through the history of philosophy. We can now approach those questions with new techniques and new findings in the brain sciences. In this course, we'll draw on both the resources of philosophical history and contemporary psychology and neuroscience in order to explore the multifaceted relationships between minds and bodies—between consciousness and the brain.

We'll trace philosophical approaches to bodies and minds from the ancient Greeks through the Middle Ages, the Enlightenment, and Renaissance to the birth of psychology and 20th-century debates. Materialism, idealism, dualism, and a range of other positions are examined in contemporary forms as well as historical contexts. Lectures include contributions from the Presocratics, Plato, Aristotle, Augustine, Aquinas, Descartes, Leibniz, Spinoza, Hobbes, Locke, Hume, Kant, Sigmund Freud, William James, Ludwig Wittgenstein, and Alan M. Turing, always with an eye to their importance from the perspective of the 21st century. Hinduism, Buddhism, meditation, and yoga are included as alternative approaches from the East, together with a history of the soul from the Orphic mysteries through the Judeo-Christian tradition.

How exactly does the brain produce the phenomena of consciousness? Drawing on today's neurosciences, we will explore the many ways that memory works; the nature of emotion, self-consciousness, and our sense of self; the routes of perception, dreams, hallucinations; and questions of free will. The course includes a tour of the brain through strange forms of consciousness: blindsight, face blindness, motion blindness, Tourette's, left-side neglect, split brains, locked-in syndrome, loss of ability to remember, and inability to forget.

Experiments on using the body to shape the mind, perception, consciousness, imagination, and memory are included for the student to perform during lectures. Scientific work on neural correlates of consciousness forms a major theme, drawing on new technologies for brain imaging, the binding theories of Francis Crick, computational theories of mind, and breakthroughs in anesthesiology.

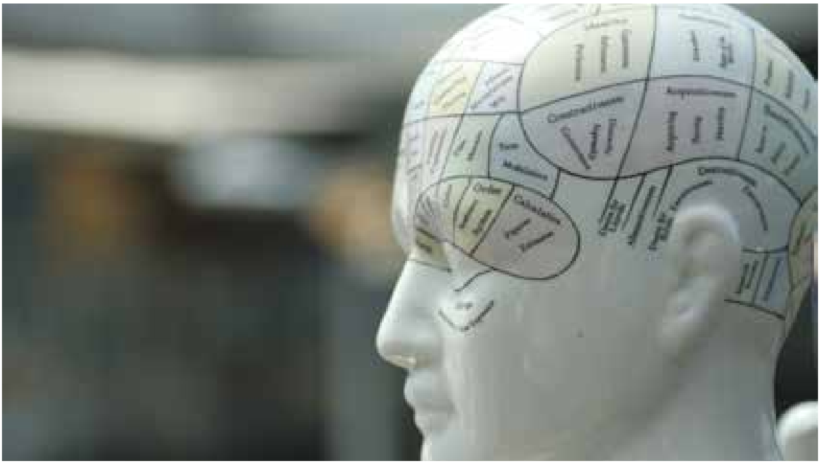
Are the core questions scientific questions or philosophical ones? Thought experiments have been called the melodies of philosophy, and are used throughout the lectures to emphasize perennial challenges to reductive materialism and functionalism. We'll also take a full tour of the lively philosophical debates now going on between monism, dualism, materialism, and its alternatives in the work of David Chalmers, Daniel Dennett, John Searle, Frank Jackson, and many others.

The course draws in depth on contemporary science and contemporary philosophy of mind. The final lecture outlines an approach to understanding body and mind through a combination of philosophy and science: a philosophical science of consciousness.

Lecture 1

Mind, Body, and Questions of Consciousness

A major question regarding bodies and minds is this: How do our physical brains produce our subjective experience? Inside a human skull is about three pounds of gray matter. It has distinct layers and an unfathomably complex tangle of specialized cells. But in the end, it's just three pounds of matter. Yet, we daily encounter the phenomena of our subjective experience: the way the sun looks, the sound of the birds, and the smell of the coffee. How can three pounds of matter produce not merely the objective phenomena of electro-chemical impulses across synapses but also the subjective phenomena of sights, sounds, touch, and smells? The standard name for that central question is the mind-body problem, and we'll be exploring it throughout this course.



Science and Philosophy

- Today's brain scientists are relative newcomers to the mind-body problem. Only over the last few decades have we started to have tools adequate to track experimentally the organizational structure of the brain. What we've learned from empirical research over even that short period is breathtaking. But we still have far to go.

- Philosophers, on the other hand, have been dealing with the central issues for centuries. Their tools have been different: the abstract tools of logic and philosophical argument.

- Brain scientists focus on the empirical details as subjective experience emerges from a physical substrate. Philosophers focus on how such a thing is even logically possible.

The Field

- When starting a journey, it often helps to have a rough map of the territory. A strategy from work by the philosopher John Haugeland is useful here: envisioning the territory as a baseball field.

- In the ballpark of mind-body ideas, materialists are in right field. For them, the only world we have is a physical world. Everything real must therefore be a part of that world. To the extent that subjective phenomena are real, they must ultimately be physical.

- The idealists are in left field. They say the only experience of the world we have is subjective experience. To the extent that anything objective is real, it must somehow be a construct from the subjective. Idealists are few and far between these days.

- The dualists play center field. For them, the world is not purely physical, as the materialists would have it. The world is not essentially subjective, as the idealists would have it. The world is composed of two different realms, neither of which can be denied: mind and body, objective and subjective. The dualists try to cover both fields.

- Today, nearly everyone in the research domain—philosophers and scientists alike—play toward the right. Nearly all researchers are materialists of one form or another. We have just one universe, a physical universe. The problem is to figure out how things that aren't objective—subjective touch sensations, tastes, and smells—can be part of that materialistic picture.

- Outfielders can play deep or shallow, and that is part of the baseball field image too. Deep in right field are the reductive materialists. As they see it, because the universe is ultimately the universe of physics, everything—both people and their subjective lives—must reduce to elements of physics.

- Not all materialists play deep. Some suggest that reality must be understood on different levels. Maybe there are explanations appropriate at the level of psychology that don't deny the ultimately physical character of the universe, but that don't reduce to physics, either.

- Regarding the idealists in left field: Some play deep and some play shallow. Deep in left field are those who think that there is no physical universe. An objectively real world is an illusion. Somehow the whole thing is made of subjectivity.

- As we move closer in, we change the question slightly: Instead of asking, "Is the universe merely a subjective construct?" we ask, "To what extent is the world as we know it a subjective construct rather than an objective reality?"

- The perennial problem for right-field materialism is how subjective experience can fit in a purely material world. The perennial problem for deep-left-field idealism is how you can start with a subjectivity that is entirely us and arrive at an objective something independent of us. In center field, dualism attempts to catch the best of both worlds.

- Playing shallow or deep in center field means taking one or another approach to the problem of interaction. Toward the infield are those who ultimately think the two realms are really one, seen perhaps from two sides. Farther back you find those who think there are two essentially different realms but that there is some point at which they make contact.

- Deep in center field are those who think, if there is any contact at all, it is a contact that will be conceptually impossible for us to see. These contemporary philosophers, called mysterians, think that the connection between the subjective and the objective will forever remain a mystery.

- If that baseball field map covers the conceptual possibilities, the truth must lie in there somewhere. Somewhere in left, right, or center field, deep or shallow, is the sweet spot that is the truth about brains and minds. But how do we find it?

Losing Limbs

- Today's brain scientists are relative newcomers to the mind-body problem. But they've become major players. It's safe to say that we have learned more about the brain in the last 30 years than in the previous 3,000. It may be time to reshape our view of mind and body in light of new information.

- One example: When an arm or leg is lost, a person may continue to "feel" the missing arm or leg. The British war hero Lord Horatio Nelson, whose right arm was amputated after he took a musket ball in 1781, was a classic case. He reported feeling his missing right hand as clenching uncontrollably, the fingernails digging painfully into the palm.

- The term "phantom limb" was coined in the context of massive numbers of amputations in the carnage of the American Civil War, when the phenomenon became all too familiar.

- It's interesting to note that Nelson himself took his experience with a phantom limb as evidence for the existence of a sentient soul existing beyond death. If the sensations of an arm can exist when it is no longer there, why not the sensations of an entire body?

- Of course, that's not the scientific explanation for the phantom limb phenomenon. For decades, the primary theory was that what the person was feeling was stimulation in the severed nerve ends, known as peripheral neuromas.

- There is a newer and very different theory of phantom limbs, which turns crucially on brain anatomy. A canyon in the brain, the central sulcus, separates the frontal lobe in front of it from the parietal lobe behind it. Along the front side of it is the primary motor cortex, from which signals output to the muscles of the body. Along the rear of the central sulcus is the somatosensory cortex, in which sensations from the body are processed.

- Sensations from particular areas of your body are processed in very specific areas of the somatosensory cortex. The somatosensory body map can reorganize when there are parts of the body from which stimulation is no longer received.

- The neuroscientist V. S. Ramachandran wondered whether the impulses interpreted as coming from a phantom hand were really coming from its severed nerves. Maybe they arose from other parts of the body that had taken over the areas of the somatosensory cortex previously associated with the hand.

- Ramachandran tested his theory with a young man who had lost an arm in an automobile accident. The subject said he could still feel the missing hand. Ramachandran stimulated other parts of the young man's body to see what produced a sensation in the phantom hand.

- He was able to map the phantom hand on two other body areas: the young man's face and his shoulder. Stimulation of those areas was read as sensation from the missing hand.

- If that Ramachandran's idea was right, some sensory input that will keep the motor impulse under control would be required. Ramachandran developed a technique to supply visual input using a simple mirror box. The amputee's remaining arm goes in a box, which creates a mirror image where the phantom arm should be. Unclenching the good arm looks like relaxing the phantom arm.

- The result of the therapy is that the phantom arm can feel that it has relaxed. Thirty-five years after an accident that left her with a phantom hand, a 57-year old woman's phantom hand pain had become unendurable. But 30-minute sessions with a mirror box over two weeks were sufficient to relieve that pain.

- It seems that the mind can create the body. The lesson of phantom limbs is that it can create a part of the body that isn't there at all.

Gaining Limbs

- The first part of the story of phantom limbs is the story of limbs lost. The second part is about limbs gained.

- Wouldn't it be wonderful if we could replace lost limbs with mechanical replacements that functioned like real limbs? DARPA—the Defense Advanced Research Projects Agency—continues to spend billions on research with that goal.

- Robotic arms have been developed with 10 to 17 miniature motors, enabling near-natural replication of arm and hand movements. A further goal is to take the artificiality out of artificial replacements: to develop a replacement arm that can be used without thinking in the same way you use your arm without thinking.

In the first 10 years of the 21st century, there were nearly 6,000 amputations in the U.S. Armed Forces. Developing replacement limbs is a huge need.

- Brain-computer interfaces represent a major step toward that goal. A crucial early achievement was muscle reinnervation, which routes severed nerves at the site of an amputation to some living muscle that is still intact.

- Motor impulses along the original nerves then produce muscle contractions somewhere else. In arm replacements, the nerves are often routed to the pectoral muscles. We can build a device that reads those muscle contractions and produces computer-controlled movements in a robotic arm.

- This process requires extensive training. But the amazing takeaway is that the brain can learn to control a prosthetic limb as its own. The mind can build a new body.

- Brain-computer interface has been used in still more extreme cases. Quadriplegia involves paralysis of all the limbs and torso. It is often caused by spinal injury in an accident, but can also be the result of multiple sclerosis or amyotrophic lateral sclerosis (ALS, or Lou Gehrig's disease).

- In both quadriplegia and ALS, the brain is still fully functional. People can still imagine moving their hands or legs. The goal is therefore to use microelectrode arrays planted directly into the skull to record electrical activity. That electrical activity is translated through complex algorithms in the computer interface to guide an artificial limb.

- Developmental obstacles are daunting, but all evidence is that the mind can learn to use an artificial body even in cases as extreme as these. DARPA reports that an individual can learn to produce a few movements of a robotic limb within two weeks and many more within five.

- Very recently, the technique has been extended even further. Brain-controlled electrical impulses have been sent to a quadriplegic's own arm and hand, activating their own muscles with electrical impulses. Bypassing the damaged neural pathway with a very indirect route, the mind has found its own body again.

Suggested Reading

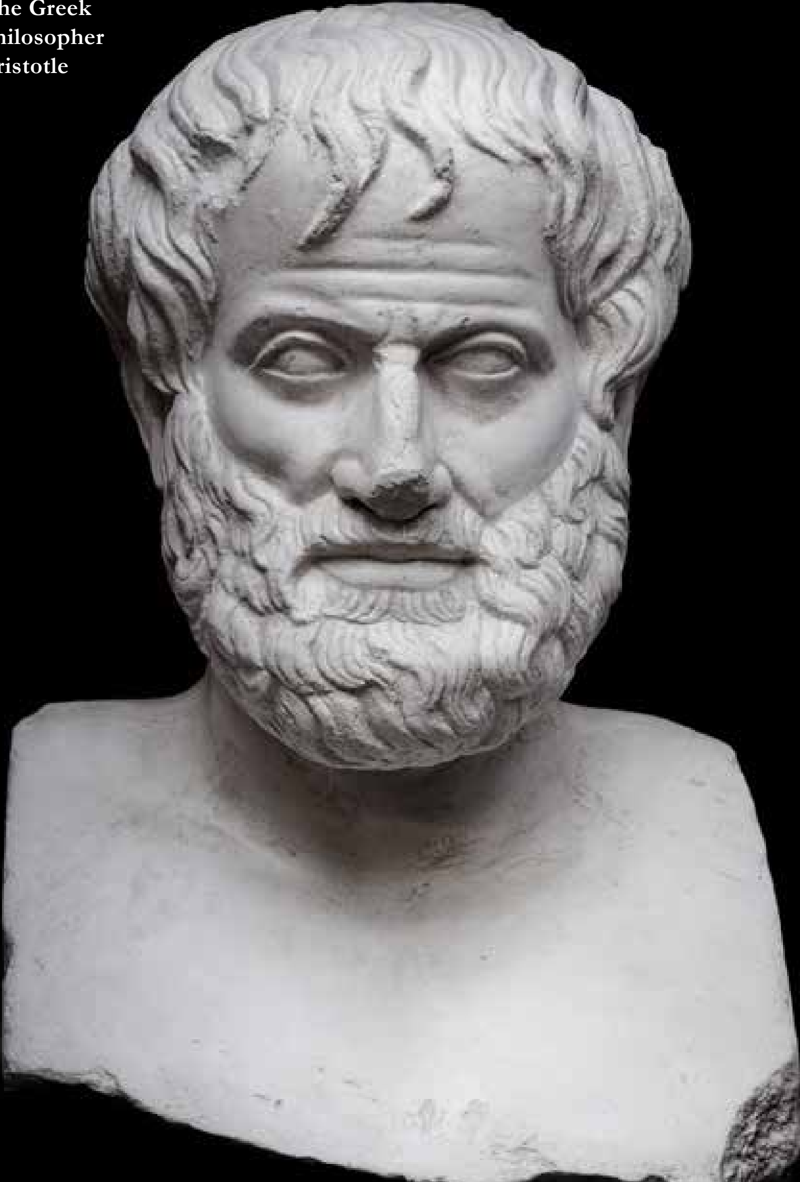
Gutenplan, ed., *A Companion to the Philosophy of Mind*.

Ramachandran and Blakeslee, *Phantoms in the Brain*.

Questions to Consider

1. If the baseball diamond is a map of all conceptual possibilities, a fly ball hit just right would land on the truth. If you had to make a guess right now, where do you think it would land? What position would you play?

2. This lecture emphasized ways in which a mind can create a body. Can interaction go in the other direction as well? Are there cases in your experience in which the body shaped the mind?



The Greek
philosopher
Aristotle

Lecture 2

Mind and Body in Greek Philosophy

This lecture discusses some of the earliest philosophical influences on thinking about mind and body. These early influences continue to resonate with contemporary work—both philosophical and scientific. The previous lecture laid out major philosophical positions on mind and body using the metaphor of a baseball field. Early forms of many of those positions can be found in ancient Greek philosophy, which this lecture focuses on. Our philosophical legacy from the Greeks includes materialism, dualism, and another force called functionalism.

The famous Greek philosopher Plato

The Beginning

- The origins of what we think of as Greek philosophy lie not in Greece but in what was known as Ionia, on the west coast of present-day Turkey. In order to find the city of origin for Western philosophy, you have to travel 200 miles east of Athens, across the Aegean sea, to the city of Miletus. Ionia was the cradle of Greek philosophy, and Miletus was the cradle of Ionian thought.

- The watershed figure in ancient philosophy is Socrates. Just as our traditional dating system runs B.C. and A.D., Greek philosophy is divided into the period before Socrates and after. Starting with the Ionians, the earliest philosophers we're talking about are those before Socrates, the Presocratics.

- The Presocratics were materialists. For them, the universe was fundamentally a material universe. With a few qualifications, Greek philosophy begins by exploring various forms of materialism.

- The central question for the Presocratics was the ultimate nature of the cosmos. The Ionians were cosmologists, asking: What is the world made of? They wanted a unified answer: a single substance of which everything else is composed.

- Histories of Western philosophy typically start with Thales of Miletus, writing around 600 B.C. He believed the cosmos was made of water. The Presocratics that followed Thales picked a different substance but offered a similarly materialistic cosmology.

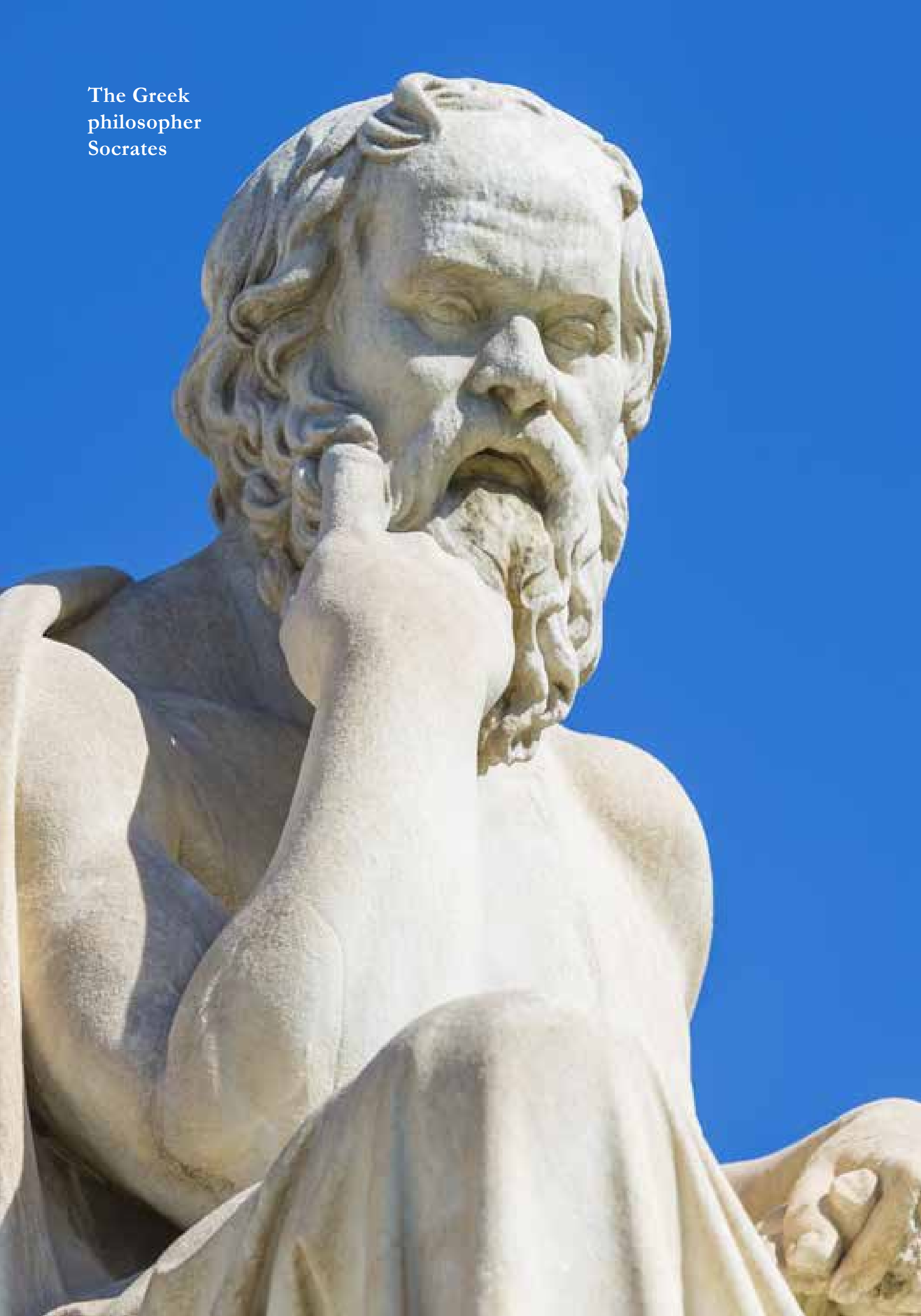
- Around 550 B.C., Anaximenes proposed that air was the fundamental substance. In Heraclitus, about 500 B.C., it was fire. Whichever matter is chosen as ultimate, we still have a firmly materialistic universe.

- Around 400 B.C., the atomists—Leucippus and Democritus—came closer to the world that is reflected in contemporary science. Leucippus and Democritus envisaged the cosmos as composed of extremely small particles moving randomly in a void, much as we think of atoms.

- Another angle came from the Pythagoreans, who flourished from the time of Pythagoras himself in the 500s B.C. through intellectual descendants into the 300s. The Pythagoreans rejected a cosmos made of water, air, fire, or any other material thing. For them, the cosmos is made of numbers.

- Also apparent in Presocratic philosophy are elements of panpsychism—a view that reappears in contemporary philosophy. Panpsychists insist that the world is made of one basic kind of stuff, but speculate that basic stuff isn't purely physical.

- In sum, the major message of the Presocratics is materialism. But you can also find mentions of the transmigration of souls and hints of panpsychism


The Greek philosopher Socrates

Socrates and Plato

- All Greek philosophy is dated before or after the watershed figure of Socrates, who lived from about 470 to 400 B.C. Socrates was a citizen of Athens, to the extent that it's right to think of Greek philosophy in terms of Athens.

- Given his importance, it may come as a surprise that we have no writings of Socrates. The simple reason is that Socrates himself wrote nothing; he believed the written word erodes memory. What we know comes from the character of Socrates in the plays written by his student Plato between 400 and 350 B.C. These were the dialogues.

- If the Presocratics give us materialism, Plato gives us dualism. The universe is composed of not one realm but two: both the physical realm and a second realm. That second realm is the realm of both ideas and of the soul.

- Plato's clearest outline of dualism is in the *Phaedo*. In that work, Socrates has been found guilty of impiety and corrupting the youth. He is sentenced to death by drinking hemlock. Socrates speaks of the soul as trapped in the body, liberated to fuller wisdom after death.

- For Plato, the real world is the world of ideas. In that regard, he is something of an idealist. Over the door to Plato's Academy was a sign that read "Let no-one ignorant of geometry enter here."

- A triangle drawn on a blackboard has with imperfect angles and bending lines. Geometry is about the ideal triangle represented by that imperfect sketch. The ideal triangle is defined in terms of three perfect angles and three perfect lines.

- That ideal triangle is the real one for Plato. The perfect triangle is too good for this world. It exists only in the realm of forms, of ideas. For Plato, what goes for triangles goes for all concepts: Any act of goodness only imperfectly represents goodness itself. Plato's world is a world divided between the realm of the ideal and its imperfect physical imitation. Plato says that when we really know something, it is the ideal that we grasp.

Immortality

- In the *Phaedo*, Plato portrays Socrates as offering four arguments for the immortality of the soul. The central argument turns on this: For Socrates, knowing demands contact with the ideal world of forms. He maintains that "learning" is actually remembering from our previous existence among the forms. He concludes that there must therefore be a life beyond the present one: a life among the forms.

- He also argues from opposites: If death comes from life, then life must come from death. Like Pythagoras before him, he too invokes the transmigration of souls.

Greek mathematician and philosopher Pythagoras

Modularity

- Modularity is a theme from Plato that is echoed in contemporary scientific work on mind and body. Our contemporary understanding of both mind and brain is in terms of modules: Parts of your brain do different things, corresponding to different aspects of your mental life. Some parts handle visual data, others handle audio data, and so on.

- The idea that both mind and brain function in terms of distinct modules has been a guiding principle throughout the history of the brain sciences. But the core idea of distinct mental modules is a legacy of Greek philosophy, appearing with particular strength in Plato.

- Plato envisaged a tripartite soul: three distinct modules of mental life. One module is the module of desire: drives of hunger, thirst, and sex. A second module is the module of *thumos*—a force of courage shown in battle. The faculty of reason is the third and should rule over the other two.

The Mind

- There is another image of mind and body anticipated in Plato that appears in contemporary research as well. This theme is captured in the title of a book by Marvin Minsky: *The Society of Mind*, in which he tries to explain how minds work.

- Minsky says the mind is what the brain does. But how? Minsky tries to show how the mind can emerge from the collaborative organization of smaller and essentially mindless bits—what he calls agents. A mind is an emergent phenomenon from the interaction of bits and pieces, just as society is an emergent phenomenon from the interaction of people.

- The analogy between minds and societies the core of another of Plato's dialogues, the *Republic*. In Plato, the analogy functions in the other direction: It is societies that are compared to minds, rather than minds that are compared to societies. In particular, Plato is interested in leveraging an understanding of how a mind works well in order to understand how a society will work well.

- Plato's soul modules come into play: The producers in his ideal society form the economic base of the state, which he identifies with the mental module of desire. Those who protect the state—the warrior class—correspond to the module of thumos. But the state must be under the charge of a rational ruling class: Plato's rational guardians.

Functionalism

- Functionalism is a view that attempts to go beyond materialism and dualism. It too is anticipated in Greek philosophy, in the work of Aristotle.

- Envisaging mind-body ideas on the image of a baseball field, idealism would be in left field and materialism in right. Dualism would be in center field trying to capture the best of both. But that image doesn't capture everything. Laid out that way, our positions all seem to be answering a certain kind of question: What is the basic stuff of the universe?

- The functionalist says that is the wrong question. Functionalists believe that we won't understand mental phenomena by looking for some specific stuff they're made of. It's a pattern and dynamics of organization that we should be looking for, not some "stuff" that the mind is made of.

- Functionalism is the view that the appropriate level of analysis is at the organization level. For example, in order to understand the mental life of an organism, we have to remember that it is an organism, with both a complex internal organization and a complex interaction with its environment. In order to understand elements of mental life—desires, beliefs, thoughts, even pains—we have to understand the part they play in the life of the whole organism.

- Aristotle can easily be seen as the first functionalist. Plato's forms existed in an ideal world apart: the abstract realm of "Plato's heaven." Aristotle was a devoted student of Plato for 20 years. But after Plato's death, Aristotle began to bring Plato's forms down to earth. The forms of things, Aristotle says, don't exist separately in some idealistic realm. They are the forms of real material things.

- In his *De Anima*—*On the Soul*—Aristotle says that the soul is the form of a natural, organized human body. If you want to find the soul, you have to look at the living organism as a whole. The soul is the form—the organization—of that living, acting, perceiving, and reasoning animal.

- Plato looked at the body and saw a cage that traps the soul. Aristotle looked at the body and saw the soul in how it thrives and how it functions.

- One implication of the shift from Plato to Aristotle is that Plato's arguments for life after death disappear. Aristotle compares the relation between matter and form, between body and soul, to the relation between a candle and its shape. When the candle is gone, its shape is gone too. When the body dies, its Aristotelian form—its Aristotelian soul—will go with it. For Aristotle, there is no life after death.

Suggested Reading

Kirk and Raven, *The Presocratic Philosophers*.

Plato, *Phaedo*.

———, *Republic*.

Questions to Consider

1. Plato says that we really understand geometry only when we grasp the ideal form of a triangle. Do you think that is right? Do you think it generalizes to knowledge in general?

2. Imagine that you were a philosopher in ancient Greece. You find yourself pressed between the materialism of the Presocratics, the dualism of Plato, or the functionalism of Aristotle. Which would you choose?

Lecture 3

Eastern Perspectives on Mind and Body

At the core of Western philosophy is the mind-body problem: If the world is composed of two radically different kinds of things—the mental and the physical—how can they possibly interact? The problem doesn't appear in the same way in Eastern philosophy at all. It's not merely that Eastern philosophy offers a different answer. In important ways, Eastern philosophy doesn't even ask the same question. Hinduism, the focus of the first half of this lecture, comes the closest. The lecture also discusses Buddhism, which is farther away from the Western tradition.

Hinduism

- The sacred texts of Hinduism are the ancient Indian Vedas, dating from 1500 to 500 B.C. There are many schools in the philosophical development of Hinduism. This lecture will concentrate on the Samkhya and Yoga schools of thought.

- Both are explicitly dualist. The world is seen as dividing into two major components. As in Western philosophy, that division is crucial to the role of mind.

- Yogic and Samkhya Hinduism cut their dualism at a different place than the Western tradition does. Although the Hindu universe is composed of two basic

realms, they aren't what we think of as the realms of mind and body, or the physical and the mental.

- On one side of the Hindu divide is prakriti. Prakriti is the realm of the natural world. Prakriti includes observable inner psychological processes—like hunger and touch—as well.

- Atman is on the other side of the divide. If prakriti is the known, atman is the knower. Prakriti includes the subjective experience, precisely because we can experience it. Atman is not the experience, but the experiencer.

- The fundamental mystery for the Western dualist is the existence of conscious experience in a physical world. The fundamental mystery for a Hindu dualist is the existence of Atman in the natural world of prakriti: the existence of an observer in a world that shows only the observed.

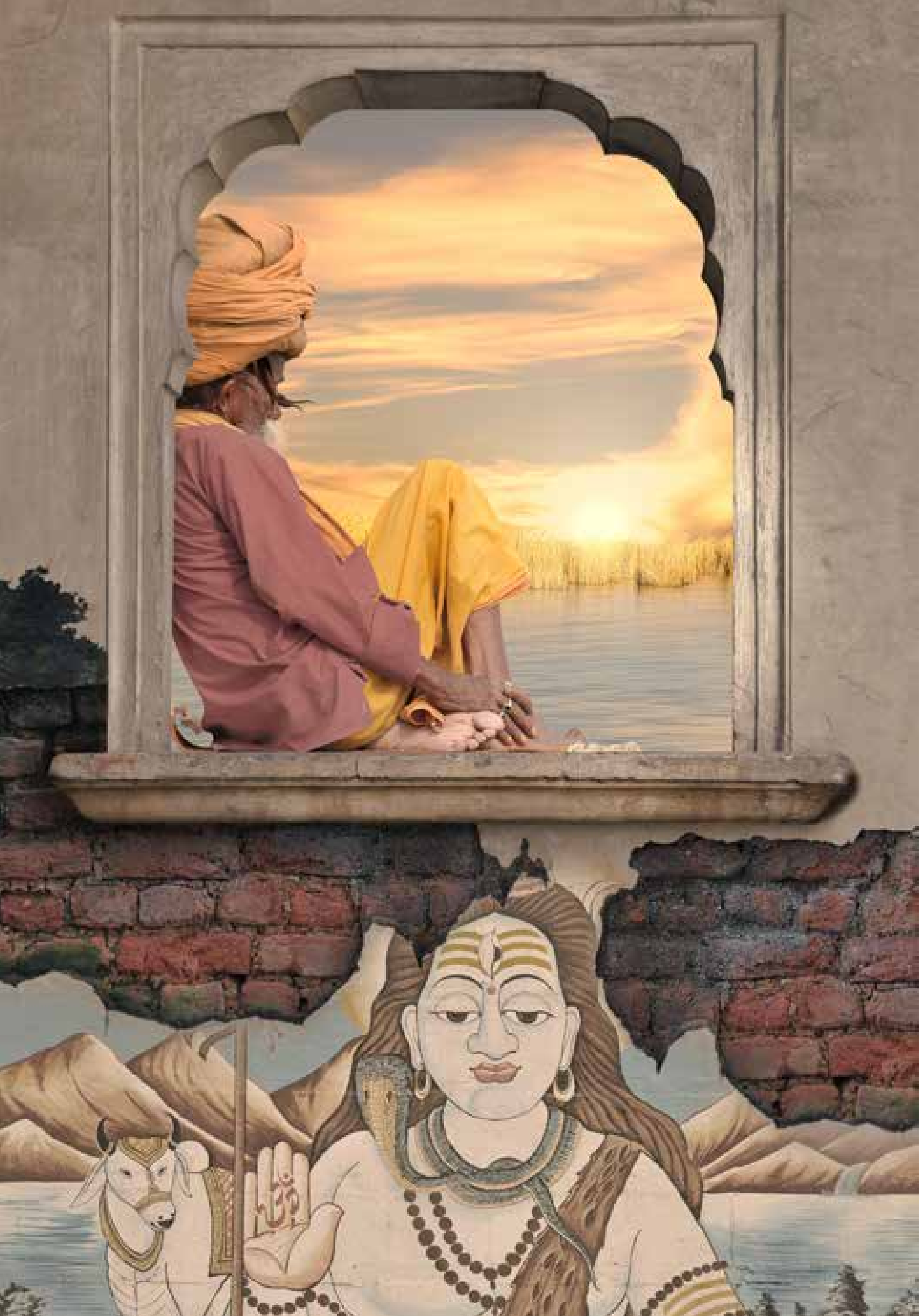
The Meaning of Philosophy

- There is a major difference in emphasis in what philosophy means in the Eastern and Western traditions. As a result, how one should approach a philosophical problem is different. What would count as a solution is also different.

- Western philosophy is first and foremost a theoretical discipline. It's an attempt at a representation and explanation of the way things are.

- By contrast, Eastern philosophy is not first and foremost a theoretical discipline. It is quite fundamentally a practical discipline. If Western philosophy is first and foremost an attempt to find out how things are, Eastern philosophy is first and foremost an attempt to figure out how to live.

- Western philosophers typically try to figure out how things stand, going from there to an inference as to how we ought to live. For example, one can figure out whether the ethical thing to do is to create the greatest good for the greatest number, and then put that theoretical principle into practice in one's own life.



- By contrast, Eastern philosophers tend to speak of truth as something that one should approach through practice rather than theory. Understanding is something one finds through practice: A bicycle rider doesn't study the physics of balance first; they understand balance by getting on and riding.

Suffering

- The practical problem of suffering is the focus of Eastern philosophy. Here, it helps to contrast Western thoughts on suffering with Eastern ones. An example from Western philosophy: In David Hume's *Dialogues Concerning Natural Religion*, the character Philo lists life's horrors as part of a theoretical argument—an argument that there can be no omnipotent, omniscient, and morally perfect God.

- The Western philosopher wants to know how evil can fit into a universe with an omnipotent, omniscient, and morally perfect creator. By contrast, the Eastern philosopher's concern with suffering is not first and foremost theoretical. The Eastern philosopher's concern with suffering is practical: How can we live with it? How can we perhaps alleviate it?

- In Hinduism, suffering is inherently observable, so it is inherently part of prakriti. Atman is not part of the world of prakriti, so it is not part of the world of suffering. If one is ruled by one's atman—if one is enlightened—one can free oneself from the pushes and pulls of prakriti. By realizing and being ruled by one's atman, one can free oneself from unnecessary suffering.

Buddhism

- The major Eastern philosophy is Buddhism. Buddhism and Hinduism share an emphasis on a practical rather than theoretical problem. The problem is the same: the problem of suffering. But Buddhism offers a very different path, a path that is not dualistic. In that sense, Buddhism offers a path even farther from the Western tradition.

- Buddhism traces back to the teachings of Siddhartha Gautama, who was born in what is now Nepal. Though the exact dates of his life are controversial, it appears that he died a few decades before or after 400 B.C. As a point of reference, Socrates died by drinking hemlock at almost precisely the same time: 399 B.C.

- Our earliest biographies of Siddhartha come from a much later period, in the 2nd century A.D. They are probably most important not as historical accounts but as representations of central values. With variations, the story is the same in many of those texts. Siddhartha is born as a prince, the son of a wealthy warrior nobleman.

- At his birth, it is foretold that he will be a great man, but that he will develop in one of two directions. Either he will become a great king or he will renounce the material world to become a holy man. The latter will occur only if he learns about old age, sickness, death, and the example of asceticism.

- Desiring that his son become a great king, his father hides him from life's harsh realities within the palace walls. No old or sick are allowed within. Siddhartha is shielded both from the fact of death and from exposure to religion. Into his 20s, through marriage and the birth of a son, Siddhartha leads a sanitized life of artificial perfection, cloistered within the palace walls.

- In his late 20s, Siddhartha journeys outside the palace walls for the first time. He is exposed to the realities of human suffering for the first time. When he first sees a very old man, his charioteer explains that all people grow old.

- Then he sees a diseased man and a decaying corpse and learns of sickness and death. He also sees an ascetic. His charioteer explains that the ascetic has renounced the world in order to escape the fear of sickness, age, and death.

- Siddhartha now knows suffering to be the human condition. Following what was foretold, he leaves his palace and his wife and son, and he devotes himself to finding a solution to suffering. Only when he finds that path does he become the Buddha, which means "the enlightened one."

- Just as in Hinduism, the problem Buddhism attempts to resolve is the problem of suffering. But the attempt to solve that problem is very different, as is the worldview that results.

Buddhism on Suffering

- A major theme in modern Buddhism is the denial of any central self. Indeed, the concept of such a self is held to be a major source of suffering. For that reason, much of modern Buddhism is anatman—without atman.

- Although experience is vivid and various, there is no observer distinct from that experience. There is no self. Underneath is merely soullessness and emptiness, or sunyata. Buddha did believe in reincarnation, but where Hinduism has many gods, Buddhism has none.

- At the core of Buddhism are the Four Noble Truths, reputed to come from the Buddha's first teaching:
    1. Suffering is everywhere, permeating not only human existence but existence in general.
    2. All forms of suffering stem from craving: thirst, greed, and desire.
    3. The cessation of suffering comes by letting that craving go.
    4. The Eightfold Path is the way to achieve the cessation of suffering. It includes right understanding, right thoughts, right speech, right action, right effort, right mindfulness, and meditation.

- The epistemology of Buddhism is arch-empiricism: What we know is what we experience. Buddhist metaphysics goes along with that: What is real is what is experienced.

- It is because of this empiricism that the Buddhist sees the self as an illusion. In Hinduism, atman must be something different—something dualistically apart—because it is by definition found nowhere in experience. In Buddhism, what is real is what is experienced, so there is no atman. With the disappearance of self, we gain the disappearance of self-centered craving. With the disappearance of craving comes the release from suffering.

- In one Buddhist parable, a house is on fire; this is the Buddha's metaphor for the suffering that surrounds us. The important thing is not to analyze the ongoing damage. The important thing is to get out of the house. Release from suffering—the practical issue—is more important than metaphysics.

- What does the world look like if we follow the Buddhist path? The result is not a dualist world. A dualism of mind and body, the mental and the physical, lies deep within Western philosophy. A different dualism—of prakriti and atman—is essential to Hinduism. But the forms of Buddhism described in this lecture deny the existence of atman. With one stroke they leave Dualism behind.

East and West

- This lecture has outlined the alternative that Eastern approaches present to a Western philosophical and scientific worldview. Regarding the historical interface between the two, it wasn't merely the East that influenced the West. The West also influenced the East.

- The empire established by Alexander the Great extended far beyond the Mediterranean, far to the east. It brought Greek influences to India and Bactria, east of the Caspian Sea, and to Buddhism in the 300s B.C.

- There are clear similarities to Greek Skepticism in Buddhism's distrust of the world of experience as fleeting, passing, unreliable, and impermanent. The dignity of asceticism appears in both Buddhism and Greek Cynicism. Most strongly, the search for equanimity by the control of one's own craving—focusing on one's own reaction to experience rather than the experience itself—is precisely the core of Greek Stoicism.

- Bidirectional influence wasn't just through Alexander's empire. The Silk Road began in the 200s B.C. as a 4,000-mile trade route from China through India to the Mediterranean. Greek philosophy, Hinduism, and Buddhism would have met and mingled along the Silk Road as well, each influencing the other.

Suggested Reading

Levine, *The Positive Psychology of Buddhism and Yoga*.

Nauriyal, Drummond, and Lal, *Buddhist Thought and Applied Psychological Research*.

Walpola, *What the Buddha Taught*.

Questions to Consider

1. Western philosophers typically try to figure out the theoretical truth, going from there to practical applications. Eastern philosophers tend to speak of truth as something you approach through practice rather than theory. Are there moments in your life in which you've taken the Western approach? The Eastern approach?

2. The core of anatman Buddhism is a denial that there is a self. But even proponents admit that is a hard mental state to sustain. How close can you come? Can you conceive of your experience without a you?

Lecture 4

Using the Body to Shape the Mind

When we think about mind and body, we inevitably think of the mind as being in charge. In this lecture, we will begin to explore what happens if we stand that image on its head. In what ways is the body in control? How might we use the body to shape the mind? This lecture explores some traditions from Eastern philosophy, which has much to teach us about using the body to shape the mind. It also covers some modern scientific findings.

Yoga

- Because of its myriad blended elements, creating a tidy intellectual history of Eastern philosophy is difficult. The blended history is nowhere more evident than in trying to track Eastern traditions of yoga and meditation.

- The origins of yoga are lost in time. The term is found in the oldest text in any Indo-European language, the Rigveda, composed in India in perhaps 1200 B.C. It is also prominent in the Upanishads, from perhaps 500 B.C. We tend to think of yoga as a physical practice, but in the historical texts, the term often represents spiritual claims as well.

- The yoga with which most of us are familiar is Hatha yoga. It is a yoga of postures and relaxation. Another form is Raja Yoga, which stresses meditation. But there is no sharp distinction between yoga as a physical practice and yoga as meditation.

- In a classic image, the mind is compared to a lake. When the surface of the lake is agitated, it becomes opaque. One cannot see within. Yogic postures, with simple movements from one posture to another, are intended to help calm the surface of the lake. The postures are called asanas. Muscle relaxation and the coordination of breath and movement are important.

- A good yoga instructor will repeatedly tell their charges to stretch the body taut in a pose but not beyond. The goal is not ego-driven achievement of some demonstrated pose. The goal is a state in which one is calmly immersed, mind and body, in breath and movement.

Meditation

- Some aspects of Buddhist meditation are only a shade away from yoga. Anapanasati is a meditation technique common to a wide range of Buddhist traditions. It's sometimes called mindfulness of breathing.

- Meditation has many forms, but it gained a major foothold in the West through Maharishi Mahesh Yogi and the Transcendental Meditation movement, also known as TM.

- The core of TM is mental repetition of a mantra. Mental repetition of the mantra offers a focus. One frees the mind by letting all other thoughts go.

- That's the core of meditative practice in all its forms: The idea is to let go of the thoughts, emotions, and anxieties that agitate the surface of the mind. Calm the lake by letting go of distracting and disturbing influences. Notice that the idea isn't to push or force those influences away. The idea is to let them go.

- The attempt to deliberately push thoughts away seems to be counterproductive. For instance, the psychologist Daniel Wegner is known for his work on the "white bear phenomenon." Try not to think of a white bear: Once the idea is introduced, it's almost impossible to get it out of the mind. In the same way, unpleasant memories and images and emotions tied to unwanted behaviors can themselves become the focus of the desire to avoid them.

Positive Effects?

- The psychologist Martin Levine has reported his own experience with yoga. His doctor had told him that his heart palpitations and aches and pains were inevitable aspects of aging. But after a month of yoga, he writes, "all my ailments—the aches and sudden pains, the heart palpitations—disappeared. I felt as if the movements had oiled my joints, so that I now moved smoothly and easily." The doctor had been wrong. As Levine puts it, "I hadn't been aging; I had been rusting."

- We have good scientific evidence that yoga does increase flexibility and balance in both young people and seniors. The physical effect for which we have the best positive evidence is relief of chronic low back pain, in both the short and long term.

- What of mental effects? Can yoga and meditation be used to shape the mind? Yoga and meditation certainly seem to produce an altered mental state. But it's not clear precisely what that altered mental state is.

- Using functional MRI, investigators have reported patterns of brain activity indicative of mind wandering are less pronounced in meditators. These differences have been found in meditators even when they're not meditating.

- As early as the 1950s, researchers hauled cumbersome EEG equipment to remote Indian monasteries in order to make measurements on Indian yogis during meditation. Those measurements indicated brain waves in the 8–12 hertz alpha range, which is characteristic of relaxed waking. The meditative state among Indian yogis also seemed to be characterized by sensory withdrawal: a state of calm that was not interrupted by noises or distractions from outside.

- Later studies with Japanese Zen practitioners showed brain waves slowing to 6 or 7 hertz, but with different results regarding external sensation. Normally the brain pays reduced attention as a sound is repeated many times. Zen practitioners responded to every repetition of the same sound as if it were new.

- The claims made for positive effects of meditation and yoga range from the very plausible to the more extreme. It is widely accepted that yoga and meditation reduce stress. Other aspects of mind can quite plausibly come with stress reduction: lower levels of anxiety and improved mood, for example.

- Claims that yoga and meditation lower blood pressure and reduce the risk of heart disease and stroke are more extreme, but those might occur as a byproduct of stress reduction as well. Still more extreme are claims regarding control of chronic pain, strengthening of the immune system, and decreasing inflammation and symptoms of arthritis and fibromyalgia. It has even been claimed the meditation can cure diabetes and cancer.

- The effects claimed at the stress-reduction end seem very plausible. Those at the cancer-curing end seem dubious. In truth, all such claims turn out to be amazingly difficult to test.

- Can yoga and meditation be recommended as ways to reduce stress? Absolutely. But we can't say that we have solid scientific evidence that they reduce stress more effectively than alternatives like simple relaxation, listening to music, or spending quality time with a pet.

- Here is an example that makes clear both the tantalizing possibility of positive results from meditation and the need for solid research: Tonya Jacobs, at the UC Davis Center for Mind and Brain, led a team that studied the impact of meditation on telomerase, which is something like an anti-aging enzyme that protects telomeres—sequences at the end of each strand of DNA.

- Thirty people were chosen to participate in a three-month meditation retreat. Thirty other people were chosen as a control group, with a similar distribution of age, sex, and body mass index. At the end of three months, the blood of each group was tested. The result was that levels of telomerase were higher in the meditation group.

- That's a wonderfully promising result. Unfortunately, the experiment isn't designed to give us the full comparison that we need. It compared results in 30 people who got to go on a meditation retreat for three months with results for 30 people who were simply told they were on a wait list. That doesn't tell us whether it was meditation or merely three months off that produced the telomerase result.

Western Thought

- Using the body to shape the mind also appears frequently in Western thought, primarily in recommendations for physical exercise.
    - Plato claimed that movement and methodical physical exercise save and preserve the good condition of every human being.

    - Thoreau emphasized the mental effects of exercise: "How vain it is to sit down to write when you have not stood up to live!"

    - Although remembered as a poet, Walt Whitman supported himself as a journalist. He wrote a 13-part series for a New York newspaper entitled "Manly Health and Training, with Off-Hand Hints Toward Their Conditions." Whitman advocates getting up early, taking a walk, and breathing the fresh air.

- Using mice as test subjects, researchers at the Salk Institute at the University of California in San Diego found compelling results regarding the growth of new neurons in the brain as a result of exercise. Mice raised in an environment that includes exercise wheels, toys, and opportunities for play grow about twice as many new brain cells as do mice in an impoverished environment.

- Charles Hillman, at the University of Illinois, has devoted much of his research career to the effects of exercise in children. He and his colleagues asked a group of children to visit their lab on two occasions.
    - On one visit the kids did 20 minutes of treadmill walking before the visit. On another visit, they rested for 20 minutes before entering the lab.

    - The exercise visit produced higher performance on cognitive tests emphasizing attention. Neural activity from frontal and parietal areas that are correlated with attention also measurably increased with exercise.

- In adults, short bouts of physical exercise have been found to enhance attention and short-term memory. Working with Ben Sibley, a colleague in physical education, the psychologist Sian Beilock found that those with poorer working memory benefited most from those short workouts.

- Exercise also seems of value later in life. One of the effects of age is a shrinking of the hippocampus. That is the module deep in the brain, important for converting short-term to long-term memory, which showed the greatest growth in exercising mice.

- In older adults, the hippocampus typically shrinks by 5 percent or more each passing decade. Kirk Erickson and his colleagues at the University of Pittsburgh found that a regular stretching routine didn't affect the rate of hippocampus shrinkage. But they found that older adults who walked 40 minutes around a track three times a week increased the hippocampus as much 2 percent a year.

QWERTY

- An example of a more subtle form of body-mind influence comes from the standard QWERTY keyboard, so named because the first six letters on the upper left are QWERTY.
- An experiment shows people two columns of letter pairs. On the left is a column with the letter pairs FV, VR, TF, JY, MJ, and UH. On the right is a list with the letter pairs CJ, GK, KT, CM, EJ, and JD.
- When asked which they like better, a majority of people pick the column on the right, starting with CJ, GK, and KT. They have no idea why.
- The interesting fact about that column is that those letter pairs are easier to type on the familiar QWERTY keyboard. People use a different hand for each letter of the pair. The other column of letter pairs is harder, requiring a person to type two letters with the same hand. This is an example of the body shaping the mind.

Suggested Reading

Beilock, *How the Body Shapes the Mind*.

Desikachar, *The Heart of Yoga*.

Questions to Consider

1. In a classic image used in both yoga and meditation, the mind is compared to a lake. When it is agitated, it becomes opaque: You cannot see within. Try the simple breathing exercise outlined in the video or audio lecture. Does that help to calm the surface of the lake?

2. In exploring whether yoga and meditation reduce stress, the lecture emphasizes the research question: Compared to what? For what kinds of questions will comparison be important? For what kinds of concerns might it not be important?

Lecture 5

History of the Soul

Despite its immense importance over the millennia, there is one concept not found in contemporary science or philosophy of mind: the concept of the soul. That's interesting because the concept of the soul is of great importance in the history of Western philosophy, from its earliest origins to the medieval period. Why would it vanish? What happened to it? Why are problems of mind and body still with us, both philosophically and scientifically, while contemporary scientists and philosophers have nothing to say about souls? This lecture considers the history of that vanished concept.

Early Origins of the Soul

- Our notion of the soul arises not from one ancient source but from several. To trace its history is something like tracing the tangled fibers of a braided rope.

- The word *soul* traces to related terms in Saxon and Old High German, Norse, and Lithuanian. It appears in Old English in the epic Anglo-Saxon poem *Beowulf*, written sometime between 700 and 1000.

- The Greek term we translate as "soul" is *psyche*. Intriguingly, even the Greek concept of psyche radically changed meaning over time. Psyche in a

Platonic dialogue of 350 B.C. means something quite different than it does 350 years earlier in Homer's epic poems, the *Iliad* and the *Odyssey*. In Homer, *psyche* means "breath." It's something that leaves a person when they die, but isn't that person.

- In the *Phaedo*, Plato's Socrates speaks of the immortality of the soul. He speaks of the soul as trapped in the body. Both of those sound far more familiar to a modern Western mind than did Homer's concept.

- Regarding another source, the Bible: Surprisingly, our concept of the soul doesn't appear in the Old Testament any more clearly than it does in Homer. The term in the Old Testament that gets translated as "soul" is *nephesh*. But according to scholars, the word *nephesh* is never used to mean a spiritual part of a person. *Nephesh* identifies a whole living being, not a part of the living being that is somehow distinct from the body.

Braiding Histories Together

- According to tradition, in 300 B.C. in Alexandria, the Egyptian ruler Ptolemy II assembled six scholars from each of the 12 tribes of Israel. He had them translate the first five books of the Bible from Hebrew into Greek. The historical translations are known as the Septuagint, after the legendary 70 or so scholars who did the work.

- In translating the first five books of the Old Testament into Greek, the Jewish scholars had to translate the Hebrew term *nephesh* into Greek. They used the closest Greek term they could find. That's where the conceptual histories join: *nephesh* was translated as *psyche*. It is in the New Testament psyche rather than the Old Testament *nephesh* that we start to see familiar elements of the modern Western concept of soul.

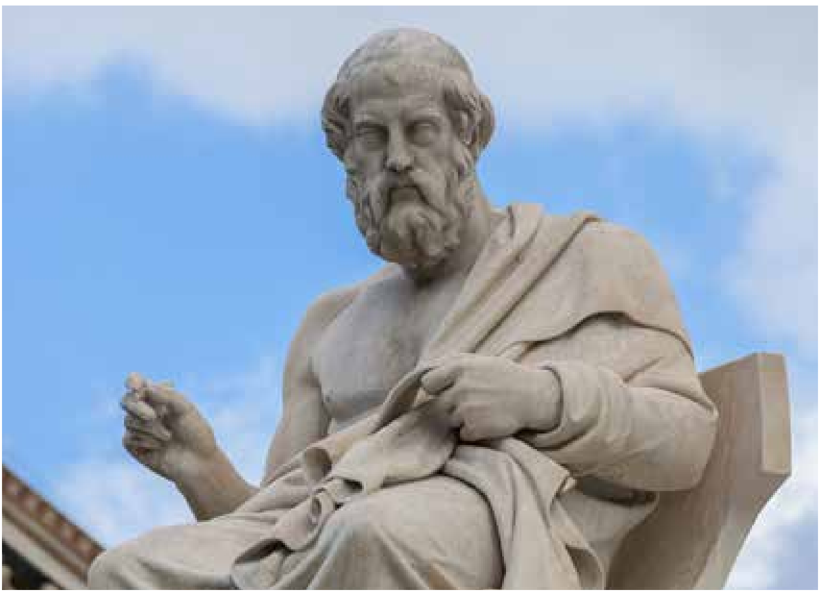
- Importantly, the promise of eternal life is a central idea in the New Testament. Surprisingly, though, even with the New Testament promise of eternal life, it's difficult to find explicit mention of a soul that leaves the body and continues at death.

Eastern Influences

- The point in history where we first found a concept of the soul that was close to ours was in Plato's dialogues of 350–400 B.C. Plato's conception owes a debt to the Presocratics, particularly the Pythagoreans.

- The Pythagoreans had a conception of the transmigration of souls—an element that also appears in Plato's *Phaedo*. Transmigration demands a concept of something that is the same soul, existing before and after its instantiation in its current body.

- If the history of our concept of soul stretches from Plato back to the Pythagoreans, where did they get their concept of soul, reincarnated from one body to another? The possibility of an Eastern influence by way of Egypt is extremely tempting. Perhaps our concept of soul traces back to Plato, from Plato back to the Pythagoreans, and from the Pythagoreans back to influences from the East.



Greek philosopher Plato

- What were those influences? Among the prime candidates are the Orphic mysteries. There is much that we don't know about the cult of Orpheus, but it seems to have come into Greece around 600 B.C. At the core of the Orphic mysteries is the myth of the god Dionysus, who like the Egyptian god Osiris was dismembered and then reborn.

- Within that mythic structure, people are envisaged as composed of a soul of divine origin trapped within a baser body. The soul is judged after death. It passes either to punishment by continuing a cycle of transmigration or to the reward of release.

After Plato

- How does the conceptual history of the soul develop after Plato? The soul in its contemporary sense really came into its own as Greek philosophy was joined with the Judeo-Christian tradition.

- In the spring of 1947, so the story goes, a Bedouin boy searching for a lost goat discovered a cave in the cliffs above the Dead Sea. Those caves contained jars filled with manuscripts now known as the Dead Sea Scrolls.

- They are attributed to a separatist Jewish sect known as the Essenes, who would have been contemporaries of Jesus. A Greek influence is already evident in the writings of the Essenes, including elements reminiscent of both the Orphic mysteries and Pythagoreanism.

- Philo of Alexandria was a Jewish thinker who lived from 25 BC to 40 A.D. He too was a contemporary of Jesus. Philo claims that the Greek philosophers must have read the Jewish scriptures. Strongly influenced by Plato, Philo maintains that soul and body are two distinct elements in man, with the soul linked to the rational elements and the body to the sensual. Liberation comes with the freeing of the soul from the body.

- After Philo, we can trace a string of religious thinkers crucial to the development of Christianity as we know it. They were all deeply influenced by the Greek philosophy that came before them, particularly Plato's work.

Saint Augustine

- There is evidence of familiarity with Greek philosophy in the New Testament letters of Paul, who was both a Jew and an educated Roman citizen. The Christian apologist Justin Martyr, writing around 150 A.D., agrees with Philo in claiming that the wisdom of Plato was borrowed from Moses and the Old Testament prophets.

- The Greek Christian philosopher Athenagoras defended Christianity before the emperor Marcus Aurelius in 170 A.D. by citing Plato, Aristotle, and the Stoics. About the same time, Clement of Alexandria speaks of Plato as Moses writing in Greek.

Medieval Philosophy

- The further development of the concept of the soul offers a clear insight into the two major figures in medieval philosophy: Augustine of Hippo and Thomas Aquinas. Their views correspond with those of Plato and Aristotle, respectively.

- Writing in 400 A.D., Augustine professes, "The question of the soul troubles many people, and I confess that I am among them." The soul, he says is "a special substance, endowed with reason, adapted to rule the body." Augustine explicitly says, "I myself am my soul."

- Writing eight centuries later in 1250 A.D., Aquinas says equally explicitly, "I am not my soul." His view is Aristotelian, rather than Platonic. The soul is not something trapped inside a body. The soul is the form of a living being, just as the body is the matter of a living being.

- Eventually, it was Augustine's approach that won out. It's not the Aristotelian but the Platonic view that gets braided with the Judeo-Christian tradition to give us the contemporary Western concept of the soul.

The Soul Today

- That history gives us a clearer sense of where our idea of the soul comes from. But why isn't it discussed in the contemporary field of psychology? There are two major figures responsible for the disappearance of the soul, one writing in the philosophy of the 17th century and one in the birth of a science of psychology as the 19th century gave way to the 20th.

- The crucial figure in 17th-century philosophy is Descartes. Descartes makes it clear that he considers the mind and soul interchangeable. In his work, the soul has been subsumed into the concept of mind.

- The crucial figure in exiling the soul from the science of psychology is William James. In his major work, 1890's *The Principles of Psychology*, James makes it clear that this is to be a science of the mind. James says that the science of psychology doesn't need the soul. His main objection is that it is superfluous. What the science of psychology seeks to understand is the actual subjective phenomena of consciousness—an empirical realm open and obvious before us.

- For James, the soul is a metaphysically and scientifically unnecessary add-on: an idle wheel that turns nothing. According to him, "As psychologists, we need not be metaphysical at all. The phenomena are enough, the passing thought itself is the only verifiable thinker, and its empirical connection with the brain-process is the ultimate known law."

- But James was also a pragmatist, who thought that we ultimately have to understand truth in terms of belief. He adds that anyone "who finds any comfort in the idea of the Soul is, however, perfectly free to continue to believe in it; for our reasonings have not established the non-existence of the Soul; they have only proved its superfluity for scientific purposes."

- In 1901, however, there was an attempt to prove the existence of the soul scientifically. Dr. Duncan MacDougall weighed six patients dying of tuberculosis, reporting a weight loss of 21 grams at the moment of death—perhaps the soul leaving the body. MacDougall also weighed 15 dogs at the moment of death, recording no weight loss. He concluded that dogs have no souls.

- Unfortunately for those seeking scientific confirmation of a soul, the experiment and its reporting were badly flawed. As his fellow physician Augustus Clarke pointed out, at death the lungs no longer cool the blood and so there is a sudden rise in body temperature.

- The resulting water loss due to a rise in sweating could easily explain the 21-gram weight loss in humans. Why doesn't that happen with dogs? Dogs have no sweat glands.

Suggested Reading

Goeetz and Taliaferro, *A Brief History of the Soul*.

Plato, *Phaedo*.

Wise, Abegg, and Cook, eds., *The Dead Sea Scrolls*.

Questions to Consider

1 What would you give as a definition of the term *soul*? On that definition, do people have souls?

2 James doesn't say there aren't souls. He says the hypothesis isn't necessary for scientific purposes. Are there different contexts in which the concept might be important?

René Descartes

Lecture 6

How Descartes Divided Mental from Physical

Concepts of mind and soul evolved from the Greeks through the Middle Ages, with many classical thinkers reflecting on the nature and relation of body and mind. What isn't obvious in those reflections is the contemporary form of the mind-body problem. This lecture will focus on the development of thought from the Middle Ages through the Renaissance to the Enlightenment. A clear case can be made that it is during this period in the 1600 and 1700s that the mind-body problem as we know it first appears with full force.

French philosopher René Descartes

Descartes and the Central Question

- Let's begin by analyzing the central question: If the mental and the physical are two such different realms, how could one arise from the other? There are two presuppositions involved in that question. Presupposition one is the idea that there are two radically different realms. Presupposition two is that one does arise from the other. Half of the question implies an unbridgeable gap; the other half asks how it is bridged.

- A crucial figure here is the 17th-century French mathematician and philosopher, René Descartes. Descartes is responsible for the sharp dichotomy that forms the first presupposition in our question. In Descartes, the mental and the physical are two radically different realms.

- A second big split arises as a reaction to Descartes: If the mental and the physical are two such different realms, how could one arise from the other? There are two radically different ways to try to answer that part of the question. Those very different reactions to Descartes echo through the Enlightenment and continue today.

Descartes's Goal

- Descartes is famous for the phrase "I think, therefore I am." Here's the context: Descartes was looking for absolute certainty. What can we really know?

- Descartes's *Meditations on First Philosophy* is the narrative of an intellectual journey, designed to carry the reader along the same path. Descartes's goal is certainty: What can we be absolutely certain of? Clearly nothing that we have been told, as everyone has been told contradictory things or incorrect things.

- Our senses and sensations within the body can deceive us as well. Take, for instance, phantom limbs, which we discussed in Lecture 1.

- Descartes's acid test for certainty is the hypothesis of an evil demon: Perhaps an evil demon has planted false beliefs in one's mind. For instance, maybe the demon deceives people into believing they have bodies when they actually don't. Only something that would survive the hypothesis of an immensely powerful deceiver can be accepted as absolutely certain.

- Can anything survive such a test? Descartes thinks the answer is yes. He thinks he has found a rock of sheer certainty: *Cogito ergo sum*, or, "I think, therefore I am."

- Interestingly, the central argument offered by Descartes echoes an argument in Augustine written nearly 1,000 years before. Here is part of the central passage in Descartes, with a segue from doubt to the kernel of certainty: "But there is a deceiver of supreme power and cunning who is deliberately and constantly deceiving me. In that case, I too undoubtedly exist, if he is deceiving me."

- Here is part of the passage in Augustine: “if I am deceived, by this same token I am. … I am not deceived in this knowledge that I am.” The argument is the same: I cannot doubt that I exist; if I doubt, I must exist.

The Division

- In his *Discourse on the Method*, Descartes argues, “I think, therefore I am” and concludes, “the soul by which I am what I am, is entirely distinct from body.” Descartes is quite explicit about identifying the mind with the thinking soul.

- Descartes was a major contributor to the science of his time. The metaphor of human physiology throughout the period in which Descartes lived, and the metaphor explicit in his work in anatomy, is the metaphor of the machine.

- For Descartes and his contemporaries, the idea wasn’t merely that machines could imitate animals. The idea was that animals were a kind of machine. It wasn’t merely that machines could imitate bodily processes, but that bodies were essentially mechanical.

- What Descartes draws from the *cogito* (“I think”) concept is not just a kernel of certainty but a big split: a portrait of a universe split between the physical on one side and the mental on the other.

The Bridge

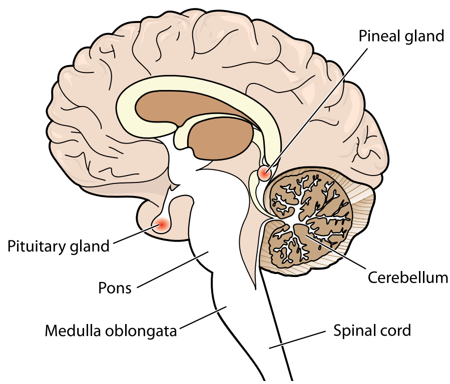
- Despite their differences, it seems clear common sense that the mental and the physical realm do interact. Take, for example, someone mentally deciding they want a cup of coffee and then physically drinking it. Or someone physically drinking Scotch and feeling a mental effect from the alcohol.

- Descartes didn’t write in an intellectual vacuum. His readers and critics were quick to see the interaction problem. They asked: If what you say is true, how could these two essentially distinct realms possibly interact?

- Descartes's dualism seems to tell us that the interaction is impossible. The interaction problem isn't just a gap in the theory. It's a built-in refutation of the theory. Fortunately, Descartes gives a response to the interaction problem in a later work.

- His answer is based on his work in anatomy, which included dissection of the brain. It's a fascinating piece of 17th-century science. Unfortunately, it's wrong.

- Near the center of the brain at the top of the spinal cord, tucked in a groove between the two hemispheres, is a tiny gland, known today as the pineal gland. Descartes proposed that mind and body interact precisely there, in the pineal gland.



- Descartes also thought, mistakenly, that the pineal gland was unique to humans. Since he thought only humans had a mind—that animals were mere machines—he had a second mark in favor of the pineal gland as the interaction point of mind and body.

- As a piece of science, Descartes's hypothesis turned out to be a complete failure. The pineal gland produces the hormone melatonin. It can be surgically removed without in any way affecting mental function, or one's sense of oneself.

- Moreover, as a piece of philosophy, the hypothesis was doomed to failure from the start. After all, the question isn't where mind and body interact. The question is how they possibly can. Given dualism, the problem isn't what or where the point of contact is between the mental and the physical. The problem is rather that any point of contact looks like a simple impossibility.

Malebranche and Leibniz

- In the decades following Descartes, other philosophers attempted to come to terms with his arguments and their implications. There arose two radically different ways to do so.

- On one side were those who accepted Descartes's arguments, but confronted the interaction problem more directly. Two attempts on that side of the split are the occasionalism of Nicolas Malebranche and the pre-established harmony of Gottfried Leibniz. Both philosophers are full-blown dualists. There are physical and mental realms, which do not interact.

- In 1674, about 25 years after Descartes's death, Malebranche proposed that physical events are never the true or direct causes of mental effects, nor are mental events the true or direct causes of physical effects. In each case of apparent interaction, there is instead an intervention by God. God is who makes a person want a cup of coffee.

- Forty years later, Leibniz offered an even more creative theory. The mental and the physical are indeed two essentially different realms. There is no real causal interaction between them.
    - But at the beginning of the creation, God wound up a physical clock that would carry forward the entire physical history of the universe. Beside it he wound up a clock of mental events to carry forward the mental history of the universe.

    - In his wisdom, God designed the clocks so that they would run perfectly in parallel, side by side, despite the lack of any real link between them. It is known as Leibniz's doctrine of pre-established harmony.

- Both of these are attempts to live with the implications implicit in Cartesian dualism. If Malebranche and Leibniz's theories seem like desperate measures, they're an indication of the desperate measures that a full-blown dualism seems to require.

Spinoza and Hobbes

- The alternative reaction to Descartes—the other side of the big split—is to see the interaction between mind and body as so obvious that it cannot be denied. Dualism, in turn, must be wrong. That view, in its various forms, is called monism. The thinking on this side of the reactive split is exemplified by the work of two other philosophers: Baruch Spinoza and Thomas Hobbes.

- Proceeding from axiomatic definitions in the manner of Euclidean geometry, Spinoza attempted to prove that the universe is composed of only one thing. What is that thing? Spinoza is explicit: "God, or Nature." For him, the two are the same. Individual things, whether they are mice or mountains, are merely aspects of that.

- The fact that spirit pervades the whole universe marks Spinoza as a pantheist. He also thinks that all aspects of the reality are logically necessary. That's why he thinks he can deduce all of reality from basic axioms by logic alone.

- Spinoza's monism is probably the high point of rationalism: the view that the key to knowledge is not experience based, on the model of scientific experiment, but pure reason based, on the model of mathematics.

- Hobbes offers a different monism in response to Descartes. His response to Descartes is clearest in his masterwork *Leviathan*, which presents a materialistic monism. He says that all sensation is motion in the brain.

- Like dualism, monism faces a basic problem: If there is just one universe, why does it seem to have such different sides? How can the objective somehow be the subjective? How can the physical somehow be mental?

Descartes in the 21st Century

- Cartesian dualism has an enormous legacy. Our everyday common sense seems to be written in dualistic terms: We think that people have bodies, that they have minds, and that the two are not the same.

- What of Descartes's "I think, therefore I am?" Bertrand Russell argues that the most Descartes can conclude is not that "I exist" but that "thinking exists." What one experiences is the thinking, not something else.

- A similar point arises in Hume's work and in Buddhism. The experiencer isn't itself evident to experience, and isn't evident to Descartes.

- The reactive split between dualism and monism continues. The dominant view in the development of the brain sciences through the 20th and 21st centuries has clearly been materialistic monism: Everything mental must be grounded in a thoroughly physical brain. Everything that is mentally real must be physically real. Much of 20th- and 21st-century philosophy of mind follows that same approach.

- But powerful challenges to materialistic monism remain. If the universe is entirely physical and is to be understood in objective scientific terms, why does an obvious part of that universe seem so different—the part immediately evident in conscious experience?

- Alternative forms of monism have found recent philosophical defenders. One line of thought goes like this: Perhaps the world itself isn't ultimately physical, or isn't ultimately physical alone. Perhaps we have to recognize the realm of subjectivity not as something that is produced by the physical, but something built into it.

- Perhaps we have to supplement our physics to recognize consciousness as a basic force in the universe as a whole. The position is called panpsychism; the idea is the mind is in everything. In it, you can hear echoes of Spinoza.

- There are also increasingly strong voices arguing for dualism. Perhaps our science inevitably leaves out the subjective half of the universe. Perhaps consciousness cannot be understood in those terms. Perhaps an understanding of consciousness demands a different kind of science—or something other than science.

Suggested Reading

Copleston, *History of Philosophy, Vol. IV*.

Descartes, *Discourse on the Method*.

———, *Meditations on First Philosophy*.

Questions to Consider

1 Both Descartes and Augustine think there is one thing that no one and nothing could deceive you about: That you are thinking, and therefore that you exist. But where does that get us? Is there any way we can build to a more complete knowledge from that?

2 The second big split in this lecture is in reaction to Descartes: the split between monists and dualists. Are you a monist—do you think the universe is composed of just one kind of fundamental stuff? Or are you a dualist—do you think the universe is divided between two basic realms?

Lecture 7

Mistakes about Our Own Consciousnesses

Nothing is more intimate and immediate to us than our own consciousness. Surely we can't be wrong about what we see, hear, taste, and feel in consciousness. It is immediately before us. We can be wrong about the external things that we think we hear: the music was live, rather than a broadcast recording. But in none of those cases are we wrong about the contents of consciousness per se. In none of those cases are we wrong about how things seem to us. Surely the core of that idea is right. Or is it? The focus of this lecture will be the ways we can be wrong, surprisingly wrong, even about our own consciousness.

An Incomplete Picture

- Consciousness is somewhat like watching a movie in a theater. There is a visual image, like the image on the screen. There is auditory input, like input from the speakers. All of our sensations come together at a point of consciousness there in the inner theater.

- The inner-theater representation may come in handy, but it won't do as a complete description of consciousness. For one thing, it leaves out some of the senses: tastes, smells, and tactile sensations are also important elements of consciousness.

- There's also a deeper conceptual problem with the metaphor of the inner theater: It only makes sense if there is a spectator. If we do have a spectator, we're doing something like hypothesizing a person in our heads, viewing the ongoing show.

- There again, we haven't explained anything: How does that inner person see the screen—with another inner screen in their own head? However easy and tempting the inner-theater metaphor for consciousness, it can't be right on its own.

Pictures and Memory

- A persistent way we characterize our own consciousness is the picture metaphor: What's it like to remember something? Like pulling an old picture from a drawer. What's it like to imagine something? Like painting a mental picture.

- But the picture comparison doesn't fit consciousness all that well. For instance, try to remember what a penny looks like. The general image is easy, but it's difficult to remember details like which way Lincoln's head is facing and what is written above his head. A true picture would make those details easy to pull up.

- People do remember in general what pennies look like. But they don't remember precise images, so remembering isn't much like a picture.

Imagination and Narratives

- Imagination and narratives are incomplete in interesting ways. For instance, picture Sherlock Holmes putting down his newspaper, rising from his chair, throwing a cape across his shoulders, grabbing his hat, and saying, "Quick, Watson, the game is afoot!"

- What color was the chair? Did he put the newspaper to the left or right? What color was the cape he threw across his shoulders? Did it have a red silk lining? There's no firm answer to those questions.

- When we really consider our own consciousness, at the very least the idea of a picture may not always be a very good fit for the experiential facts.

Color Blindness

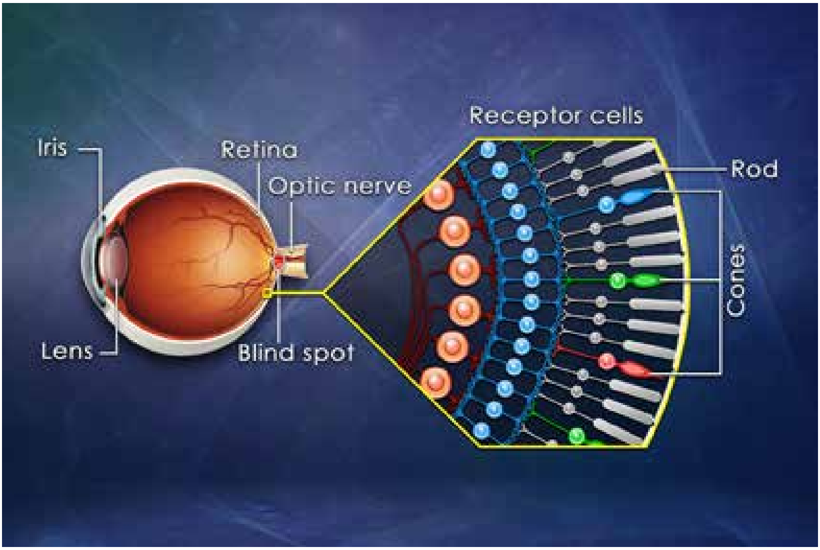
- Color blindness can teach us something regarding our own consciousness and the mistakes we make regarding our own consciousnesses. The physiological basis of the most common type of color vision is well understood. Receptor cells in the retina include rods and three types of cones.

- Rods are active in low light, sensitive to brightness; they are associated with vision in black and white. There are three varieties of cones, with peak sensitivities at short, medium, and long wavelengths. These correspond to something like the blue, green and yellow-green regions of the spectrum.

- Normal color vision is keyed not merely to the individual activation of these but the balance between them. For example, there are no true red cones: Red light is perceived because of strong stimulation of the long wavelength cones relative to the other two.

- Most color blindness is genetic. The three different types of cones operate with three different photopigments, which are encoded on the X chromosome; if something goes wrong there, color blindness can result.

- There are 6.5 to 7 million cones in each eye, concentrated toward the center of the retina. But there are 120 to 130 million rods in each eye. Cones thin out toward the edges, leaving practically nothing but rods at the edges of the retinal field. That's where everyone is color-blind: In peripheral vision, people can see dark and light, but are terrible on color.



- Here's the lesson regarding consciousness: If a person were asked to paint a picture of what their visual field looked like, knowing only what we've just learned, they might be tempted to paint a picture that was rich in color in the center, but bleached out to black and white at the edges.

- Yet in consciousness, the visual field doesn't seem to be colored in the center and black and white at the edges. How are we to explain the discrepancy between how things seem to us and how our visual field actually works?

- "Filling in" is one explanation: Our consciousness fills in colors where they don't exist, giving the impression that what we see is a full-color image all the way to the edges. Our visual field doesn't look like it's black and white at the edges because we fill in color there.

- Here's an alternative explanation: That part of the visual field just doesn't operate in terms of color. The brain doesn't expect to see colors out there. The philosopher Daniel Dennett has been particularly prominent in arguing for this approach. We don't have a brain that's busy painting in missing colors in the big picture. We have a brain that pays attention to color in central vision and ignores the lack of it at the edges.

The Blind Spot

- Blind spots are another case in which we use a picture metaphor for consciousness. Here's how to "see" the blind spot: Draw an X and a solid black dot about two inches apart on a three-by-five-inch index card.
    - Turn the card so that the X is on the right and the solid black dot on the left. Now put your hand over your right eye and hold the card at arm's length with the X straight in front of your left eye.

    - Keep focusing on that X, without moving your eye, as you bring the card slowly toward your nose. When the card is about one foot away, the dot disappears. Bring the card still closer and the dot reappears.

    - Now flip the card so that the X is on the left and the solid dot is on the right. Cover your left eye and do the same experiment again. When you get the card about a foot away, the dot will disappear again.

- The basic physiological explanation for the blind spot is as simple as the physiological explanation for color blindness. We don't have to go deeper than the retina. In order for an eye to work, impulses from the retinal cells have to transfer to the brain. The feat is accomplished by binding over a million nerve fibers into a single optic nerve.

- The bundle leaves the eye right where a blind spot is detected. The blind spot is there because the exiting optic nerve makes things too crowded for light-detecting photoreceptors. There are no rods or cones; people are totally blind in that part of the visual field. In a literal picture of what the retina actually receives, the image would have a gap in it where the blind spot is.

- Now let's take the disappearing-spot trick one step further. Take the three-by-five card and draw some cross-hatching across the side with the dot on it.

- Now repeat the experiment: Bring that card progressively closer to your eye. Here's what you'll find: The dot disappears, just as before. But it doesn't look like the cross-hatching does. The same trick works with wavy lines. The dot disappears, but the rest of the pattern remains.

- The brain has to compensate for the fact that it has no information from a certain area in the visual field. It fills it in with what it thinks is there: cross-hatching, or wavy lines, or whatever the brain has—so the explanation goes. In another respect, the brain is terrible at filling in blind spots: After all, there is a dot there.

- Daniel Dennett tries to offer an alternative to the filling-in picture for blind spots. Dennett says that the brain merely registers that there is more of the same in the area of the blind spot. It doesn't need to fill in. It merely tells you not to expect anything different there, so you don't.

Suggested Reading

Dennett, "Quining Qualia."

Hoffman, *Visual Intelligence*.

Questions to Consider

Imagine a duck dressed as a pirate. Which of these questions can you answer, without adding to what you've imagined? Is it wearing a sword? Does it have scars? What color are its feet?

1. Is its tail straight or curly?

2. One explanation for what you see in your peripheral vision is that your brain fills in color at the edges. Another explanation is that your brain simply ignores the fact that you don't see color at the edges. Which theory do you think is right?

3. One explanation for what you see in your blind spot is that your brain fills in a visual background there. Another explanation is that your brain simply tells you that there is more of the same there. Which theory do you think is right?

Lecture 8

Strange Cases of Consciousness

We can learn quite a lot about the mind and the brain by exploring our own normal consciousness. But we can also learn a great deal by looking further afield at some very strange cases of consciousness. Some strange cases of subjective experience offer important lessons as to how the brain functions. The first set of lessons from consciousness tells us about modularity in the brain. The second set of lessons tells us about multiple routes in the brain. The third set carries lessons about differentiation and separation between parts of the brain, particularly between different sides of the brain. And the fourth set of lessons concerns strange effects of blending, intersection, and union between different parts of the brain.

Modularity

- Different areas of the visual cortex have been found to process different aspects of the stream of information from the retinas, with quite different effects. The first area of information arrival is known as V1, short for visual processing area 1. The representation that appears on V1 is laid out like the image in a fish-eye mirror.

- In the adjoining area, V2, stereo vision is processed: putting together the information from two eyes to generate your impression of a three-dimensional scene. V3 is a distinct area that processes depth and distance. V4 is devoted to color, V5 to motion, and V6 is at work processing the position of objects.

- The celebrated neurologist Oliver Sacks tells the story of a successful artist, known for abstract color canvases, who suffered damage to the V4 area in an automobile accident. He lost color entirely. He spoke of the world as going not only into shades of gray, but going indescribably "wrong." What had appeared "flesh-colored" now appeared "rat-colored."

- That is a lesson regarding modularity: Those with changes to V4 have no trouble reading a newspaper, no trouble navigating the world in terms of place or motion, and no difficulty in recognizing family and friends. It is just color that is impacted.

Multiple Routes

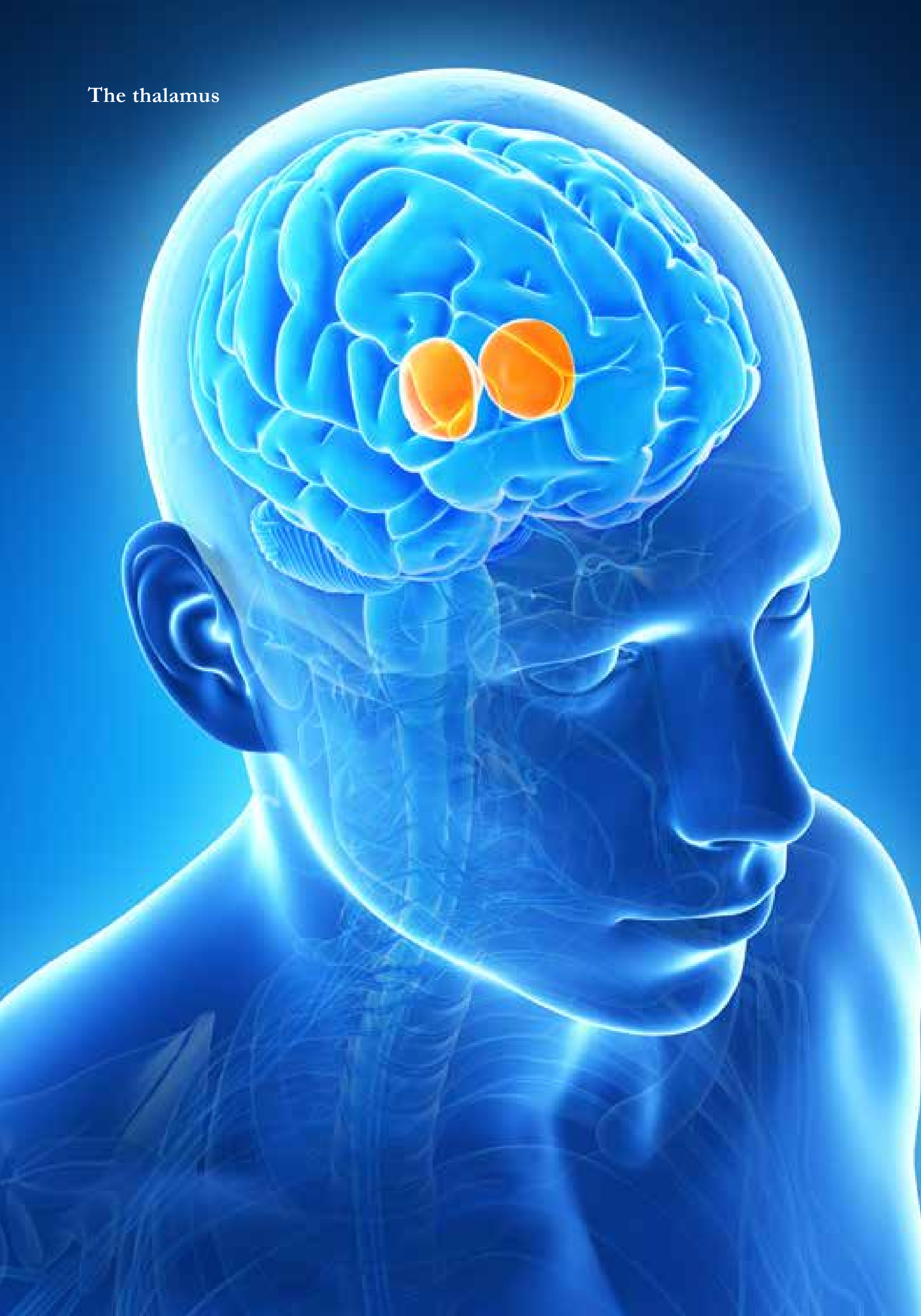
- The brain routes information through multiple paths. There is an area of the brain toward the bottom of the temporal lobes known as the fusiform gyrus. With damage to the fusiform gyrus, a person is unable to recognize faces. It's called prosopagnosia, or face-blindness.

- The more severe the condition, the more compensation is required. Those with prosopagnosia may focus on particular features as indicators of particular people. They may, in fact, prefer friends who have recognizably large noses, bushy mustaches, or the like. There are cases so extreme that people fail to recognize their own face in a mirror.

- Normal recognition of people actually involves two routes. From the fusiform gyrus—where faces are recognized—visual recognition information normally is routed to the amygdala, a very old and central part of the brain that is tied to emotion.

- But it is possible for the first route to be operative while the second is not. It is possible to have facial recognition with no accompanying emotional overtone. That strange form of consciousness is called Capgras syndrome. It results in the belief that one is surrounded by imposters; someone might think their wife is actually an imitator.

The thalamus



- The current theory is that it is that second route—the emotional one—has been interrupted. The facial recognition of the fusiform gyrus remains intact but the emotional response isn't there.

- There is an even stranger form of double routing or redundancy in the brain known as blindsight. In the anatomy of visual consciousness, information from the optic nerve comes into V1. Before that, it has been divided into left and right visual fields.

- If the V1 on one side of the brain is knocked out, one loses half the visual field. Knock out V1 on the right side and a person loses the left side of their visual field. Knock out the V2 on the left side and a person loses the right side of their visual field.

- A person whose V1 has been impacted on the right side will be unable to tell where a spot of light is on a screen to their left. They can't see over there. Their eyes may still be functional, but they are cognitively blind to the left. Strangely, however, if asked to point to where the point of light is, they may point to the spot with astounding accuracy.

- Blindsight was first documented in the work of Lawrence Weiskrantz at Oxford. Working with his original blindsight patient, Weiskrantz found that although the patient repeatedly insisted that he couldn't see to his left—a clear result of damage to the V1 area on the right—he would "guess" the position of a spot of light with 99 percent accuracy. Subjects with blindsight have been documented navigating through rooms, avoiding furniture they cannot consciously see.

- The explanation is a secondary route in the brain. When information routes from the retina, there is a "transfer station" in the thalamus: the lateral geniculate nucleus. At that point, the information stream divides. The stream goes to the normal visual processing area at the back of the head—to V1, unless of course it has been knocked out. But the stream is also routed to an evolutionary older pathway down to the brainstem in an area called the superior colliculus.

- It is that route—a route that is not revealed to consciousness—that allows accurate guesses as to the location of a spot of light. It is that route that allows blindsight navigation of a room one cannot consciously see. The brain is modular, but is also redundant.

Differentiation

- The third set of lessons about how the brain functions concerns differentiation in consciousness and the brain. The brain has two hemispheres. The right hemisphere receives sensory input from the left side of the body and controls movement on that side. The left hemisphere receives sensory input from the right side of your body and controls that side.

- The exception is the eyes. The right hemisphere receives input not from the left eye but from the left side of the visual field, taken from both eyes. The left hemisphere receives input from the right side of the visual field—again, from both eyes.

- What happens when one specific side of the brain is impacted? An example case is the neuroscientist Barbara Lipska. One day, as she stretched out her arm to turn on a computer, she realized that her right hand had disappeared. A brain tumor had knocked out her right visual field.

- There is another strange form of consciousness, parietal neglect, caused by a stroke. With this condition, people stop caring about one side of their vision. They simply neglect the left side of their world. The condition is caused by damage to the parietal region of the right brain. Interestingly, it happens only with the right brain. Lisa Genova has written a fascinating novel based on the condition, entitled *Left Neglected*.

- A person with parietal neglect will only eat food placed on the right side of the plate, ignoring even a particular treat to the left. The person may be able to see that food, and indeed may eat it once someone calls attention to it. But unprompted, they won't pay any attention to it.

- Neglect applies not only to what is seen but what is imagined. When patients in Milan were asked to imagine that they were walking through a familiar cathedral square in a particular direction, they described only what they would have seen on the right side.

- In a previous lecture, we discussed V. S. Ramachandran's astounding treatment for phantom pain using a simple mirror box. Could the same technique be used to treat parietal neglect? What if we used a mirror to show a person with parietal neglect the left side of the world?

- Ramachandran found that with the mirror in place, a person with parietal neglect clearly sees the person and the pen reflected in the mirror. But when asked to reach out and take the pen, she doesn't identify the person as to her left. Despite the fact that she has dealt with mirror images her entire life, she cannot reach out to the left, to the real person reflected, in order to grasp the pen reflected from the left. She says that the mirror is in front of the pen.

- In parietal neglect it is indeed like the left side of the world ceases to exist. When reflected in a mirror to the right, it has come into existence on the right side, but trapped behind the mirror.

The highlighted section shows the parietal lobe in the brain.

The Split

- The strongest lesson regarding differentiation between different hemispheres of the brain appears in split-brain patients. Running between the hemispheres is a broad tissue of nerves called the corpus callosum. In extreme cases of epilepsy, as a last-ditch effort, there are cases in which the corpus callosum has been surgically cut. With the exception of an overlap of functioning in the face and neck, cutting the corpus callosum effectively isolates the two hemispheres of the brain.

- There is nothing particularly remarkable in the behavior of split-brain patients. In the 1940s, a number of patients were treated for epilepsy by cutting the corpus callosum. Despite that drastic surgery the patients showed no significant ill effects. It wasn't until the 1950s, when Ronald Myers and Roger Sperry developed techniques for addressing the two hemispheres separately, that the strangeness of split-brains began to appear.

- The Myers and Sperry work began with cats and monkeys. With the corpus callosum of their animal subjects cut, they were able to train each of the hemispheres independently. What they found was that what was learned on one side of the brain did not transfer to the other side. The corpus callosum was necessary for information transfer between the hemispheres.

- Linguistic processing is almost always in the left brain. Input to the right half of the visual field—processed in the left brain—can therefore be reported verbally: If someone sees a pipe, they can say, "I see a pipe." Because linguistic processing is standardly restricted to the left brain, input to the left half of the visual field—processed in the right brain—cannot be reported verbally.

- One of the most dramatic cases involved a young man who had undergone split-brain surgery but who did retain a limited linguistic capability in his right hemisphere. Under control of his right hemisphere, this patient was able to use movable letters to spell out words with his left hand.

- He was asked what he wanted to do with his life. He responded verbally with his verbal left hemisphere that he wanted to be a draftsman. But his left hand, expressing his right brain, spelled out "automobile race." His right brain wanted to be a racecar driver.

- From his investigations, Sperry concluded that each hemisphere does seem to have its own separate and private sensations, with the right hemisphere constituting a second conscious entity running along in parallel with a dominant stream of consciousness in the left.

Synesthesia

- Synesthesia is a blending of different sensory modalities. In one form of synesthesia, music comes in colors. One note might be red, another blue. It has been estimated that synesthesia of one form or another is evident in as many as 2 people in 100.

- Synesthesia, first documented by Francis Galton in the 19th century, appears to be largely genetic. It is quite demonstrably real. The color perception of numerals can be revealed by a test in which numeral 2s and numeral 5s—computer generated in a square format so that they are mirror images of each other—are scattered in patches across a field.

- The two symbols are so similar that most people just see a field of black and white S- and Z-shapes. But for those with synesthesia who see numerals in color, patches of 2s and 5s stand out because of their different colors.

- Ramachandran's hypothesis regarding synesthesia is that it is overlap between neighboring brain areas that produces numbers with color. One of the strange facts about synesthesia is that it is seven times more common among artists, poets, and novelists. Ramachandran's further hypothesis is that overlap, blending, and hyperconductivity across various parts of the brain may be an important part of artistic creativity.

Suggested Reading

Genova, *Left Neglected*.

Nagel, "Brain Bisection and the Unity of Consciousness."

Sacks, "The Case of the Colorblind Painter."

Questions to Consider

1 People with achromatopsia live in a world without color. They can neither see nor even imagine color. Can you imagine living a month in a world without color? Can you imagine not being able to imagine color?

2 Here's an experiment to do with a friend. Give them a series of numbers, asking them to write down a color that seems right for each one:

9 1 4 7 6 2 3 5 8

In a week or so, ask them to do it again, giving them the same numbers in a different order:

7 5 2 6 3 9 1 8 4

Compare the two lists. For most people, the colors assigned the two times won't match up. If your friend has number synesthesia, they will assign the same color to the same number each time.

Lecture 9

Altered States of Consciousness

This lecture will look at altered states of consciousness from two directions: What can the brain sciences tell us about altered states? And what can altered states tell us about the brain? The lecture begins with the altered state of dreams. Then the lecture moves on the hallucinations before entering a discussion on altered states caused by drugs. Finally, the lecture looks at near-death and out-of-body experiences. As we will see, altered states offer a wealth of information about the brain.

Sleep

- There are two kinds of sleep, REM (rapid eye movement) and NREM (non-rapid eye movement). There is a standard sleep cycle that can be tracked in terms of brain waves. Waking life is characterized by beta waves, at 12 to 40 cycles per second. As we doze off into stage 1 NREM sleep, there is a shift to slower alpha and theta waves, 8–12 and 4–8 cycles per second.

- In stage 2 of NREM, rapid and rhythmic brain activity, known as sleep spindles, appear against that background. Stage 3 NREM is marked by the addition of still slower delta waves to the mix. In stage 4, deep NREM sleep, the brain settles into delta waves almost exclusively.

- Then the surprise: a rapid ascent through stages 4, 3, and 2 NREM into REM. The EEG for REM sleep actually looks a lot like waking, dominated by beta waves. But during REM sleep all input is internal, stimulated from the brain stem, rather than reaching us from outside. REM sleep lasts 20 to 30 minutes. It is followed by another descent into NREM. The entire cycle takes about 90 minutes. Eight hours of sleep represents about five cycles.

- It was long thought that dreaming was the sole province of REM sleep and that NREM sleep was empty of mental content. It now appears that there are two kinds of dreaming, one characteristic of REM and the other of NREM sleep. If you wake someone during REM sleep, they typically relate the vivid, sensation-rich, complex and bizarre dreams we are all familiar with.

- Waking someone during NREM sleep produces a different kind of report. Here dreams seem to be thought-like rather than experience-like. Sleepers will say, "I was just thinking" about something. NREM dreaming takes the form of looped fretting, almost always, of course, about things that aren't really worth fretting about.

The Purpose of Dreams

- There is no consensus on the purpose of dreams. One view is that what we do in dreams is rehearse patterns of action from the day—action repertoires, either alone or in bizarre combinations. The psychologist Nicholas Humphrey proposes that rehearsal in dreams is one of nature's most audacious and ingenious techniques for self-education.

- At the opposite extreme, Francis Crick and Graeme Mitchison proposed that dreams flood circuits at random in order to wash away unneeded connections acquired during the day. Dreaming is for forgetting rather than remembering.

- The philosopher Owen Flanagan thinks they have no real purpose all. He claims that dreams came along as a side effect as we evolved mechanisms for things that really do matter: sleep and waking consciousness.

- Do dreams reveal something deep about our psyches? Freud thought dreams were wish fulfillments. His colleague Carl Jung made a list of dream archetypes with standard meanings. Both theories have been roundly debunked.

- Nevertheless, dreams might mean something after all. A tempting theory of dreams is that during REM sleep the neural patterns that constitute images of people or events, memories, and behavioral routines are stimulated at random by impulses from the brain stem. The faculties of the higher brain then attempt to construct a narrative that makes sense of this stream of random mental contents.

- Have you ever had a dream in which you knew you were dreaming, and found you could manipulate the story in your sleep? That's a lucid dream. Fifty percent of people questioned say they've had a lucid dream at least once in their lives. Twenty percent say they have a lucid dream at least once a month.

- Sleep researchers have confirmed that lucid dreams do occur. In fact, the psychologist Stephen LaBerge runs the Lucidity Institute, and one of its aims is to teach people how to produce and control their lucid dreams.

Hallucinations

- A hallucination is defined as a sensation of something that isn't there. Although we normally think of hallucinations as visual, auditory hallucinations are the most common. There are also smell hallucinations, taste hallucinations, and touch hallucinations. Auditory hallucinations can be as simple as a hissing sound or a constant tone, and they can be teasingly soft or disruptively loud.

- The most common time for hallucinations to occur is just as one is falling asleep (hypnagogic) or waking up (hypnopompic). These hallucinations are considered normal. Hallucinations can also come with migraines, Parkinson's disease, temporal lobe seizures, lesions in the brain stem, and, of course, hallucinogenic drugs.

- What causes hallucinations? Two basic mechanisms have been proposed, though they might in fact function together.
    - The first theory emphasizes higher brain function. A person's sense of reality has slipped: The sounds and images are entirely internal, misinterpreted as coming from the outside.

    - The second theory says we are asking the wrong question. Instead of asking why we sense something that isn't there, we should be asking why we don't hear those internal sounds and see those internal images all the time. This theory says that we actually do; it's just that those internal impressions are usually swamped by input from the outside.

- If this second theory is true, hallucination should occur when input from the outside is cut. Indeed, one of the most reliable ways to produce hallucinations is by sensory deprivation. You don't need to block all the senses: Put someone in darkness for extended periods of time, and they will have visual hallucinations. Block their hearing and they will have auditory hallucinations.

- Some hallucinations, as the previously mentioned hypnagogic and hypnopompic hallucinations, are considered normal. On the other hand, hearing voices *may* occur as a symptom of mental illness. Schizophrenia involves a chronic and severe loss of contact with reality, with hallucinated voices—from God, the devil, or one's dog, for example—as a major sign.

Drug-Induced States

- Nicotine and caffeine are stimulants. So are cocaine and amphetamines. Both cocaine and amphetamines require larger and larger doses to produce their effects. Long-term use of amphetamines can lead to paranoid hallucinations. Long-term use of cocaine can produce a very specific hallucination: bugs crawling under one's skin.

- Alcohol, barbiturates, and tranquilizers are depressants. Heavy and extended use alcohol can produce hallucinations, both during use and in withdrawal.

- The altered states produced by the highly addictive narcotics heroin, morphine, codeine, and methadone are primarily of altered mood rather than perception. Altered perceptual states are first and foremost the territory of hallucinogenic: psilocybin from "magic mushrooms," mescaline from peyote, and LSD.

- Hofmann mentioned circles and spirals as part of that first acid trip. Complex hallucinations may include cartoon characters, scenes from childhood, or beautiful landscapes. But simple and specific geometric patterns show up repeatedly in visual hallucinations, whether natural or drug-induced.

- These simple patterns include wavy lines, circles, spirals, and concentric patterns.Similar patterns can be found in ancient art from around the world, some of which may be tied to rituals incorporating hallucination.

- Why the ubiquity of concentric circles and spirals? A group of mathematicians led by Jack Cowan in Ohio and Paul Bressloff in Oxford have offered a theory built on the anatomical structure of visual processing the brain: The standard patterns of base-level hallucinations could be explained by drug-induced instability within the specific brain structure of visual processing area V1.

Out of Body

- Out-of-body experiences are a state in which someone seems to see the world from a location outside their body. These may occur as an effect of hallucinogenic drugs, in association with epilepsy and migraines, or under conditions of stress. Some 15 to 20 percent of people claim to have had at least one out-of-body experience.

- The first question, of course, is whether what is perceived is real: Has some part of a person—their consciousness, perhaps—really exited their body? That would certainly have important consequences for the mind-body problem.

- Leaving one's body, projected to another place, is a common element in folklore in many cultures. "Spectral evidence" played a major role in the Salem witch trials. Witnesses testified that they had a dream or vision in which they were attacked by the accused witch. But many accounts of out-of-body experiences strain credulity.

- If out-of-body experiences literally involved leaving one's body, one should be able to perceive things that aren't visible from where one is laying. In a 1968 paper, the parapsychologist Charles Tart reported a sleep experiment in which a woman asleep in his lab was able to correctly report a five digit number hidden on a shelf above the bed.

- Was she able to see that number in an out-of-body experience? Tart's experimental controls were bad, and it seems equally plausible that she got up and peeked. Later experiments with 100 subjects who thought they could read remote messages in out-of-body experiences showed a success rate of zero. The alternative explanation is that out-of-body experiences are "as-if" experiences: the experience is merely as if you were viewing your body from the outside.

- Transcranial magnetic stimulation produces an electric current in the area of the brain under a magnetic coil. The Canadian neuroscientist Michael Persinger reports producing out-of-body experiences by transcranial magnetic stimulation in the temporal lobe.

Near Death

- A composite picture of commonly reported near-death experiences would include an out-of-body experience, travel through a tunnel toward bright light, the presence of a central figure or lost loved ones, and immersion in positive and loving emotions.

- Over the course of a year, a hospital in Southampton, England interviewed all survivors of cardiac arrest. Four percent reported near-death experiences. In the Netherlands, of 344 patients resuscitated after cardiac arrest, 18 percent reported a near-death experience.

- Interestingly, and not implausibly, many of those who have had near-death experiences regard them as life changing, and report stronger belief in an after-life and a reduced fear of death.

- Not so often reported is the rate of hellish near-death experiences: a cold field of dead gray, or cackling demons with naked people writhing in pain. The rate of these has been estimated as high as 15 percent, though people are understandably anxious to deny or forget them.

- Just as out-of-body experiences are offered as evidence of a separate soul, near-death experiences are offered as evidence of a life after death. Are they?

- The strongest form of evidence would be the occurrence of such experiences—or any experiences—within a period of flat EEG, standardly associated with bodily death. That isn't the kind of evidence we have. What we have in all cases of near-death experiences are reports after the event. It is not clear that what is reported is an experience during a period before the EEG goes flat, during a flat EEG, or even within a period of recovery.

- Reports of near-death experiences are also consistent with the dying brain hypothesis: Loss of oxygen to the brain can produce what researchers at the University of Michigan's Center for Consciousness Science call a "brainstorm" of increased electrical activity and a massive release of neurotransmitters. Visions of tunnels and lights appear with disinhibition of the visual cortex.

- Out-of-body experiences may be linked to activity in the temporal lobe. Positive and loving emotions may correlate with endorphins and enkephalins, released under stress.

- It is interesting to note that if there is a realm after death, it seems to be an all-inclusive and ecumenical realm. Christians report seeing Jesus and angels in the course of near-death experiences. Hindus report meeting the king of the dead and his messengers. If near-death experiences are a purely physical brain phenomenon, that's exactly what you'd expect: a core phenomenon interpretable in different ways within a wide range of cultures.

Suggested Reading

Flanagan, *Dreaming Souls*.

James, "On Some Hegelisms."

Sacks, *Musicophilia*.

Questions to Consider

1. The lecture mentioned three theories of dreams:

    - They serve as rehearsals of action repertoires formed during the day.
    - They flood circuits in order to wash away unneeded connections made during the day.
    - They are an evolutionary by-product that serves no function at all.

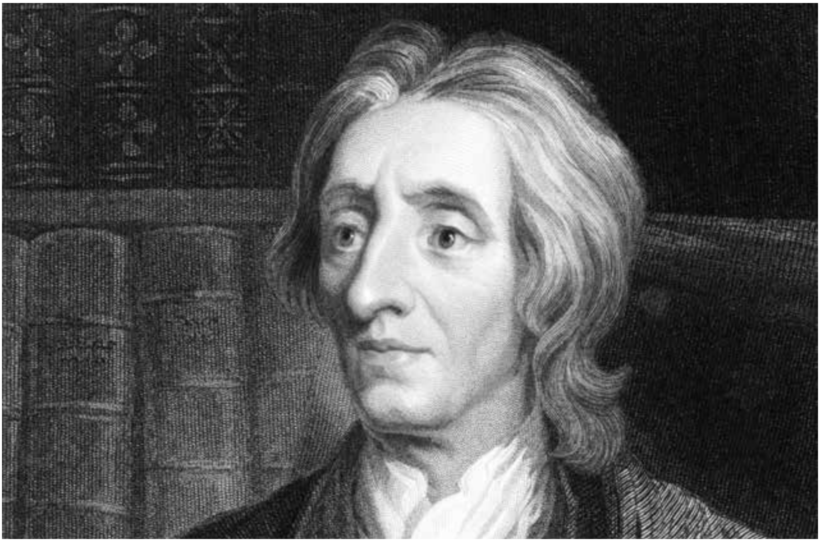
    Which theory do you think is right?

2. Have you ever had a hallucination? An out-of-body experience? What do you think we can learn from those experiences?

Lecture 10

Memory, Mind, and Brain

This lecture will begin to trace the link between memory and self, starting with a look at different types of memory and the functions they serve. In thinking about memory and the self there is one central figure in the history of philosophy: John Locke, one of the founders of British empiricism. It is in his 1690 *An Essay Concerning Human Understanding* that the term *consciousness* appears with its contemporary meaning for the first time. At the core of that work is a connection he draws between memory and personal identity: a link between memory and being the person who you are. This lecture takes a look at just what memories are and then circles back to Locke's and others' views on memory.



John Locke, who helped found British empiricism

Types of Memory

- There are many types of memory. One is short-term memory, also called working memory, which we use all the time. For example, people can remember a telephone number long enough to dial it. But at any given time, short-term memory is limited to only 15 to 30 seconds in length. Additionally, according to work by George A. Miller, we can only hold about seven pieces of short-term information, give or take a couple.

- Short-term memory occurs behind the forehead, at the front of the frontal lobe.

- In order to operate for more than 15 to 30 seconds, or with more than seven items, we need long-term memory. Examples are remembering an appointment next Tuesday or that Bill owes John 10 dollars from a lost bet last week.

- Other examples are how to shuffle a deck of cards or how to ride a bike. Remembering how to do something is called procedural memory. Procedural memory happens in a very different part of the brain, indeed, in different parts of the brain.

- When someone is first practicing a new motor skill, the frontal cortex lights up with activity. Five or six hours later, the motor skill operates more easily, much more smoothly. At that point, it's the motor cortex (looped in an arc over the center of the brain) and the cerebellum (at the back just over the spinal cord) that light up as the person goes through the routine.

- Remembering specific events—for example, when a relative stepped off the train with various details of the scene—is called episodic memory. Remembering that FDR died before the end of World War II or that you have an important appointment is called semantic memory. Both episodic and semantic memory are thought of as explicit or conscious memory.

- The hippocampus, a structure deep in the center of the brain, is crucial for laying down long-term memories of these types. Precisely how long-term memories are laid down remains an unknown. One thing we do know is that "recycling" plays a role: If the same pattern recirculates through the hippocampus, the memory is laid down with greater strength.

- We also know that sleep can play a role. The hippocampus is active during sleep. Damage to or removal of the hippocampus impacts memory in a specific way. You can still use short-term memory and may retain some long-term memories, but you become unable to lay down new long-term memories.

- In addition to facilitating the conversion of short- to long-term memories, the hippocampus codes basic spatial memory. Brain scans on humans show spatial navigation as a function of the hippocampus.

- Near the hippocampus is another important structure: the amygdala, thought of as a center for emotion. It has been suggested that the hippocampus and the amygdala function together: that emotion enhances the coding of memories.

- In evolutionary terms, the hippocampus is an ancient part of the brain. Right next to it is another ancient structure: the processing area for smell. That may partially explain why smell can be so evocative of memory.

What and Where Are Memories?

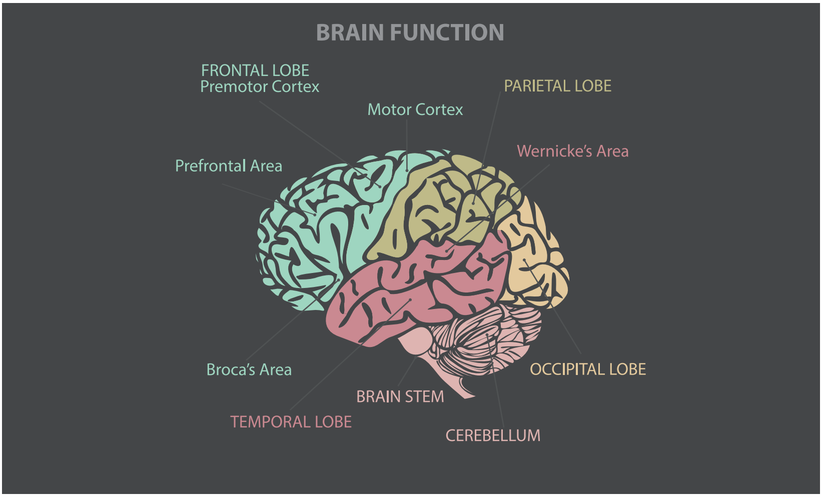
- What are memories, and where are they located? Eric Kandel won a Nobel Prize in 2000 for work addressing that question. It is not an individual neuron that encodes an individual memory. It is a pattern of neurons that encodes a memory.

- Every experience fires a particular pattern of neurons. During repeated, partially repeated, or overlapping patterns, the synapses between the neurons in that particular pattern are strengthened. Kandel's work allows us to track down the idea of neural cell assemblies in biological detail.

- Where are the neural cell assemblies stored? All evidence indicates that the patterns are stored precisely where they first appeared: Visual data is processed in the visual center, audio data in the auditory center, and so on. Perceiving those things together involved a pattern of neurons coordinated in different parts of the brain. That pattern of precisely those neurons is codified as the memory of the event in the same place or linked places as in the original perception.

- In short-term memory, things are recorded sequentially: Asked to recall the fifth number in a seven-digit sequence you just heard, you'll probably count from the start until you hit the fifth number. In contrast, memories are retrieved from long-term memory by association.

- Cell assemblies serve more than one purpose. An individual neuron may fire when a person sees a picture of the Eiffel Tower. The same neuron fires—and by supposition the entire cell assembly is activated—when a patient remembers the Eiffel Tower. That's our evidence that the neural pattern of initial perception and of later memory are the same.



- But the neuron may also fire when the person is asked merely to imagine the Eiffel Tower. The same cell assembly, in other words, may do the work of perception, memory, and imagination as well.

Memories versus Imagination

- If they memories and imagination recruit the same cell assemblies, what is the difference between memories and imagination? How can we tell them apart?

- David Hume, an empiricist in the tradition of Locke, thought that memories have an internal characteristic that allows us to tell that they are memories. Perceptions are vivid and intense. Memories are fainter, pale shadows in comparison.

- That doesn't help distinguish memories from imaginings, and not many people have been convinced. One can have a vivid memory and a tenuous piece of imagination. But one can also imagine something vividly while remembering it only vaguely.

- The truth is that Locke's problem is a problem that we all face: There may not be any internal difference between the products of memory and the products of imagination that allows us to tell them apart just by looking at them from the inside.

- Each time you bring up a memory you bring it up in a different context, and when you put it back it may not be precisely the same cell assembly. Every time you remember something your memory is liable to change. This leads to the problem of false memories: things we think we are recalling as accurate memories—accurate records of a past event or a past experience—but which are not.

- Elizabeth Loftus is an expert in the field who has done some wonderful experiments in false memory. In one, she asked parents of an advanced age to recount some events that had happened during the childhood of their adult children.
    - She then interviewed the adult children, asking them whether they remembered the events their parents had told her. In doing so, she used three episodes their parents had related.

    - But she also slipped in a fourth event that hadn't been mentioned by the parents and hadn't ever happened. About 25 percent of the adult children also insisted that the fourth fictional event had occurred to them. In some cases, they would even elaborate on remembered details of the event that never happened. They were merely imagining it.

- It turns out that our memories are extremely malleable, extremely vulnerable to suggestion. A suggested detail that is merely imagined may be incorporated as if it were part of a memory. We may later be unable to tell the genuinely remembered parts from those that were merely suggested or imagined.

- One of the gut-wrenching places this shows up is in what are termed recovered memories under therapy, including memories of being the victim of child abuse. In the 1980s, under therapy and hypnosis, cases began arising of adults who remembered being sexually abused as children.

- There have been a number of legal cases in which a perpetrator—often a father—has been publicly accused on the basis of memories thought to have been suppressed but later recovered.

- The debate regarding recovered memories continues. On the one side, child abuse is a horrible crime, and we do indeed want perpetrators to face justice. On the other side, there are cases in which the recovered memories have turned out to be false memories.

- False memories are also a frightening possibility in eyewitness testimony. Next to DNA, the testimony of an eyewitness is the gold standard in criminal cases. Yet decades of research—much of it inspired by Elizabeth Loftus—has shown that the memories recounted in eyewitness testimony are as malleable and subject to distortion as any others.

Aging

- Age does affect memory, but not as much as you might think. Perhaps because of news coverage, a large proportion of the aging population swears that they are starting to suffer from Alzheimer's or dementia. The truth is that only about 10 percent of people between 65 and 100 suffer from anything so clinical.

- The typical and normal kind of memory loss that comes with aging is the "tip-of-the-tongue" memory problem; for instance, trying to recall a name but having it remain just out of grasp. Misplacing the occasional name happens to everyone, even if it does become more frequent with age. In Alzheimer's, one loses concepts of common objects and familiarity with common places.

- True short-term memory is not much affected by normal aging. What appears to suffer most with age is the conversion of short-term memories into long-term memories. That is a function of the hippocampus, which shrinks with age. That may be why, with age, people may find it easier to remember what happened 10 years ago than what happened two months ago.

- One of the very last aspects of memory to fail—even in extreme cases of Alzheimer's and dementia—is procedural memory, remembering how.

- The good news is that memory, like other mental functions, can be preserved and extended through life-long use. Studies of people beyond the age of 70 show that those with more education have more efficient memories, with less change in memory ability. The theory is that they have learned, practiced, and can apply more flexible strategies for establishing and managing memory.

Remembering Everything

- The Argentinian author Jorge Luis Borges has a story called "Funes the Memorius," in which the title character remembers everything. Everything Funes has ever seen—every nuance of every shadow of every leaf of ivy on a wall—is etched indelibly in his memory.

- Funes is what we all would be if Locke had been right about memory: If every experience we had made an impression that was filed away in a storehouse of memory. Locke forgot forgetting. Forgetting is probably just as important for effective thinking and living as memory.

- An important point of the story is that Funes regards his prodigious memory not as a blessing but as a curse. He hides himself away in a dark room. He says, "My memory, sir, is like a garbage heap."

- The real though rare condition that Funes has is called hyperthymesia. Jill Price is credited as the first established case, with about 25 confirmed cases since.

- Price can remember every day of her life from when she was 14. Interestingly, her recall is almost entirely autobiographical: She actually performs poorly on standardized memory tests. Like Funes, she describes it as a burden: "non-stop, uncontrollable, and totally exhausting."

Suggested Reading

Borges, "Funes the Memorious."

Foer, *Moonwalking with Einstein*.

Miller, "The Magical Number Seven, Plus or Minus Two."

Questions to Consider

1. Give three examples from your own experience of each kind of memory: An episodic memory (of a scene), a semantic memory (of a fact), and a procedural memory (of how to do something).

2. What is the first thing you remember in your life? How confident are you that that is a genuine memory?

Lecture 11

Self-Consciousness and the Self

How is memory tied to our concept of the self? That's the topic this lecture explores, using John Locke as our point of reference in the history of philosophy. For Locke, it is the continuity of conscious memory that defines someone's identity as a person. But there are two distinct questions that Locke runs together as if they were the same question. Distinguishing between those two questions offers a clearer understanding. The first question: What precisely is the sense of self—is it an element added to experience, or is it built into the structure of our experience? The second question: What counts as the "same person" over time?

The Sense of Self

- Our experience seems to come with a sense of the self that is having that experience. Descartes puts the "I" at the core of his certainty: "I think, therefore I am."

- Locke says that "when we see, hear, smell, taste, feel, meditate, or will anything we know that we do so." Emphasizing a sense of continuous self, he says that it is beyond doubt that "I that write this am the same myself now whilst I write that I was Yesterday."

- Immanuel Kant, writing a little less than 100 years later, says that the "I think" accompanies all experiences, which would be impossible without it. William James says that within the stream of consciousness there is a sense of self, "felt by all men as a sort of innermost citadel within the circle, of sanctuary within the."

- The neuroscientist Antonio Damasio echoes all of these: "Besides the images of what we perceive externally," there is also "this other presence that signifies you, as observer of the things imaged, potential actor on the things imagined. If there were no such presence, how would your thoughts belong to you?"

- All of this seems obvious: When it comes to the experiential question, we do have a sense of self.

The Same Person

- What is it that makes someone the same person over time? That's the question of personal identity.

- Let's distinguish qualitative from numerical identity. Two things are numerically identical if they're literally the same entity. If two people have Subaru Outbacks, they have qualitatively similar vehicles, but not numerically identical.

- When we talk of changes in someone's personality, we use the qualitative sense: "He's not himself today" or "She's not the woman I married." We're bemoaning a qualitative change in personality while relying on the fact that we're talking about the same person in a numerical sense.

- Locke believes it is continuity of memory that makes us the same person over time—not identity of bodies. Locke's argument relies on this thought experiment: Suppose that we implant the consciousness of a prince in the body of a cobbler. We are to imagine that the prince's consciousness, including his consciousness of his past life, is implanted in the body of the cobbler. Who will that person be? Locke's answer is unequivocal: the prince. It is the continuity of conscious memory that makes a person the same person over time.

Dissociative Fugue and Multiple Personalities

- On January 17, 1887, Reverend Ansel Bourne withdrew $551 from his bank in Providence, Rhode Island, and disappeared, only to reappear in Norristown, Pennsylvania, believing he was a man named A. J. Brown. Later, he eventually awoke one morning as a very confused Ansel Bourne.

- The pioneering psychologist William James suggested they could explore the case by putting Bourne under hypnosis. Indeed, under hypnosis, A. J. Brown reappeared and was able to describe his travels to Norristown. But even the Brown that appeared under hypnosis claimed to know nothing about Reverend Ansel Bourne.

- Today the case would be classified dissociative fugue: a case of reversible amnesia for memory, personality, and personal identity.

- We've previously discussed cases in which the hemispheres of the brain are separated by surgically cutting the corpus callosum: A man might say he wants to be a draftsman but spell out that he wants to be a racecar driver with his hand. That certainly sounds like William James' conclusion: "Mr. Bourne's skull … covers two distinct personal selves."

- That was Michael Gazzaniga's initial thinking when he started working on split-brain cases. He changed his mind later because of the role of what he called the interpreter.
  - The left brain of a subject was shown a chicken claw. The right brain was shown a snow scene. The subject was asked to choose a related picture with each hand.

  - He chose a picture of a chicken with one hand and a picture of a shovel with the other. This is where the interpreter comes in. When asked, "Why did you pick two different things," the left hemisphere interpreter would make up something that covered the two responses. In this example, the subject said "Oh, that's easy. You're going to need a shovel to clean out the chicken coop."

- Because of the overarching role of the interpreter, Gazzinga ultimately decided that even in split-brain cases, there is only one dominant hemisphere, only one high-level consciousness, and in that sense only one self.

Dealing with Odd Cases

- A number of thinkers have tried to deal with both normal self-consciousness and the aforementioned unusual cases by distinguishing between different senses of self. James, for example, speaks of two selves: a "me" and an "I." The me includes all those things one tend to call mine: "my social life," "my virtues and vices," and "my memories." James clearly thought of Bourne and Brown as different selves in that sense.

- But beneath that self, James insists, there is also an "I." That is the self we feel as the focus of all our experience—the lasting core self.

- Distinctions between senses of self are also prominent in the work of neuroscientist Antonio Damasio. Damasio characterizes the core self, which is reminiscent of short-term memory, as "a transient entity, ceaselessly re-created for each and every object with which the brain interacts."

- The other self is what he terms the "autobiographical self." That self includes lasting characteristics like your name, your history, whether you tend to avoid conflict, and how you approach a problem. The autobiographical self has the properties of long-term memory.

Animals

- Are animals self-conscious? Do they have a sense of self? The standard test for sense of self in animals is the mirror test, developed by the psychologist Gordon Gallup in the 1970s. The test is this: We show the animal a mirror for the first time.

- Then, in some way the animal isn't aware of—under anesthesia, for example—we put a colored mark on its ear, or its forehead. We then watch how it then behaves in front of the mirror. Does it look at the image and then investigate the mark by reaching to that spot on its own body? If so, it must realize that the image it sees in the mirror is an image of itself. It must, therefore, have a sense of self.

- Chimpanzees pass the mirror test, as do orangutans and bonobos. Gorillas do not, except for one: Koko, a gorilla raised in close human contact and taught elements of American sign language. Human infants don't pass the test until somewhere between 18 months and two years old. There have been positive reports of versions of the mirror test for dolphins, orcas, and magpies.

- The test isn't without flaws: Many species view staring as a threat, which makes using a mirror problematic. That may complicate results. If the test does reveal a sense of self, it's unclear what specific self it reveals.

- Damasio thinks that only simple levels of autobiographical memory are present in other animals. Gordon Gallup, on the other hand, the man who designed the mirror test, thinks that chimpanzees not only recognize themselves in the mirror but have an awareness of self in terms of a personal past and future.

The Teletranslator

- Derek Parfit is a contemporary philosopher who has explored questions of memory and personal identity in great depth, using some strange and wonderful thought experiments.

- Parfit borrows the idea of a teletransporter from science fiction. Take this thought experiment: Let's say a teletransporter can take a complete map of your brain and body. Then it can disassemble your atoms and send them to a teletransporter receiver on. There, identical atoms and molecules are assembled to recreate you.

- When you step into the teletransporter, are you suddenly whisked across the universe, or do you die? An atom-by-atom replica of you appears on Alpha Centauri. But maybe it's just a Xerox copy. It's not you. You ceased to exist when your atoms were disassembled here in the teletransporter on Earth.

- Raising further doubts: What if the teletransporter sends the same set of signals to two different places in the universe, making two Xerox copies? Split-brain cases are hard enough. Parfit hands us a split-person case. What happened to the one numerical self-identical you in the process?

- Parfit thinks the thought experiments show that being the same person isn't always a matter of yes or no. Our concept gets gray and fuzzy at the hypothetical edges. Parfit calls someone who thinks that selves come in discrete yes-or-no packages an ego theorist. Parfit rejects ego theory. There just aren't individual selves in the way we intuitively think there are.

Suggested Reading

Dennett, “Where Am I?”

Locke, *An Essay Concerning Human Understanding*.

Parfit, “The Unimportance of Identity.”

Questions to Consider

1. Were Reverend Ansel Browne and A. J. Bourne the same person or not?

2. What is it that makes you numerically the same person as your past self as a child?

3. Would you step into Parfit’s teletransporter? Why or why not?

Lecture 12

Rival Psychologies of the Mind

In this course, we've been tracing the history of thought on bodies and minds, always with an eye to contemporary scientific results. Psychology, as the science of the mind, is clearly a major part of that history. There are three claimants to the title of father of psychology: William James, Sigmund Freud, and Wilhelm Wundt. All three developed their theories in the infancy of psychology as a discipline in the latter part of the 19th and early part of the 20th century. All three have influenced the trajectory of psychological theory and research to the present day. This lecture contrasts the work of James against that of Freud, and then compares both to Wundt's work.

William James

- William James's foundational work, *The Principles of Psychology*, was published in 1890. It had taken 12 years to write and was about 1,200 pages long. The publisher encouraged him to publish a shorter version, which was released in 1892 as *Psychology: the Briefer Course*.

- For James, the primary concern of psychology is the issue of consciousness. This central focus dictates both his methodology and the core of his theoretical contribution. James does not take Freud as his intellectual sparring partner in his major work. It is rather the British empiricists he is arguing against: Locke, Berkeley, and Hume.

- According to the empiricists, the mind acquires individual ideas, sensations, or impressions from sensory experience and then combines them into higher units in consciousness. For example, if we combine the idea of a horn with the idea of a horse, we come up with a unicorn.

- James sees that synthetic approach as fundamentally unscientific. He claims that a genuine science must be built on the phenomena of experience, and experiences don't come in atomic bits and pieces. The phenomena of experience always come as a whole. According to James, we have to start with what we directly know, with "total concrete states of mind."

- The psychology that James envisages is a scientific psychology based on data of a particular kind: introspective data. The target is an understanding of consciousness, and consciousness can only be revealed to us from our own subjective experience in introspection.

- Four immediate characteristics of consciousness guide James's examination.
    1. Every state of consciousness is a personal consciousness; every state of consciousness is owned by someone. With the ownership of every state of consciousness comes privacy: Every mind keeps its thoughts to itself.

    2. Within each personal consciousness, states are always changing.

    3. Consciousness is sensibly continuous. You go to sleep and wake up, but your consciousness is sensed as a continuation of what went before.

    4. Attention focuses and shifts within consciousness.

Sigmund Freud

- In an early work, Freud says that any psychological theory must meet the demands of natural science. But he also says it must explain the puzzling things that we know in consciousness.

- This is a break: James tracks consciousness as the fundamental fact, relying on introspection as the primary tool. Freud attempts to explain the puzzling things about consciousness. According to Freud, the mental is divided into three realms: the conscious, the preconscious—that which can be retrieved from memory—and the unconscious.

- Given that map of mental territory, Freud develops a theory of basic forces, outlined like a cast of mythological characters. The id is the locus of inborn biological instincts or drives. The ego, developed from the id in infancy, is the id's interface with reality. The ego attempts to bring the demands of the id into accord with the demands of the external world. The superego develops in early childhood as we incorporate internally the social constraints of parental influence.

- Freud used his theory in the treatment of patients. His writing is framed in terms of cases or generalizations from cases in which he attempts to treat hysteria, paranoia, and neuroses. The treatment protocol develops into psychoanalysis, a combination of talk therapy, free association, and dream analysis. The conceptual framework proliferates to include Oedipal complexes, Electra complexes, and so on.

- Despite Freud's insistence throughout his career that his goal was a science of psychology, his approach is closer to the therapeutic model of medicine. Freud himself resisted the use of controlled experiment. He seemed happy to see perceived success in individual cases as a vindication of broad theories.

James versus Freud

- James and Freud arrived at very different foundations for very different sciences. What did they have to say about each other?

- Despite the fact that they were contemporaries, it is very difficult to find passages in which James mentions Freud by name or the reverse. Instead, they mention an approach or position which clearly is to be taken as that of the opponent. On both sides, the mention is always is sharp and critical.

- James totally rejects the concept of an unconscious mental state. The existence of an unconscious flies in the face of any attempt to build a science of mentality grounded first and foremost in conscious introspective experience.

- James says the attempt to introduce unconscious mental states "is the sovereign means for believing what one likes in psychology, and of turning what might become a science into a tumbling-ground for whimsies." That is a strong rejection of Freudian theory.

- For his part, Freud minces no words in condemning the other side: "Whereas the psychology of consciousness never went beyond this broken sequence of events … the view that held that what is mental is itself unconscious enabled psychology to take its place as a natural science like any other."

Attacks

- Freudian psychology, in particular, came under increasing attack as pseudoscientific. A prime mover in that attack was Karl Popper. Popper wasn't primarily concerned with when a theory is true because he knew that even the best scientific theories might turn out to be wrong. He wanted to know what made a theory *scientific* to begin with, whether true or not.

- Popper's demarcation criterion was falsifiability: A theory is scientific only if it is falsifiable. Popper's cases of pseudoscience included Marxism and psychoanalysis. Those theories seem to be able to explain the results of social movements or individual psychology in every case, no matter what happens. That is the whole problem.

- A Marxist can find what he takes to be confirming evidence on every page of the newspaper. And whether a man tries to save a drowning child or drowns a child himself, the Freudian can explain it in terms of repression or some other notion. For a Freudian, every conceivable case can be explained in Freudian terms.

- Popper's falsifiability criterion has left a lasting impression. In a critique of Freud, author Richard Webster suggests that psychoanalysis may be the most complex and successful pseudoscience in history.

- However, if Freudian psychoanalysis is seen not as a single unified theory but as a collection of different theories, that critique may go too far. A number of philosophers and psychologists have come to the conclusion that at least some Freudian concepts are falsifiable, even if others may not be.

- After extensive examination, both the psychologist Hans Eysenck and the philosopher Adolf Grunbaum came to the conclusion that Popper's blanket statement was wrong. At least some aspects of Freudian theory are empirically testable. Even so, Eysenck claims that Freud managed to set back the study of psychology and psychiatry by 50 years or more.

- James's reputation has not done much better. The psychologist Gregory Kimble took a critical look at James' major work, *The Principles of Psychology*, on the 100th anniversary of its publication. He concluded that he couldn't find a single principle in it.

Wilhelm Wundt

- Although Freud and James have both been named as fathers of psychology, their theories have not fared particularly well. It is the third figure—Wilhelm Wundt, also spoken of as a father of psychology—who is a far clearer precursor of academic psychology as it developed through the 20th century.

- The object of Wundt's investigations were "the elements of consciousness." His background theory was in the tradition of the British empiricists. Wundt was looking for elements of sensation as they compounded into higher ideas.

- Unlike James's personal introspection and unlike Freud's clinical practice, Wundt's approach was experimental. Like James, Wundt took introspection to be the route into consciousness. But he was acutely aware that the attempt to focus on one's own mental state may change the mental state.

- Because of that, Wundt didn't rely on verbal introspective reports but on a subject's simple signals regarding subjective experience—the pressing of a key when a sound was heard, or when a light reached a particular point on the screen. He tested reaction times and discrimination thresholds. His experimental contexts called for stimuli that could be strictly controlled and strictly repeated with a number of subjects.

- In Wundt's work, unlike that of James and Freud, one has real psychological experiments, many of which are indistinguishable in format from psychological experiments performed today.

- In the course of the 20th century, at least in academic psychology, the influence of both James and Freud was swamped by experimental psychology in the tradition of Wundt. During his career, Wundt trained 116 graduate students, including many Americans. It was at precisely this time that psychology became popular, with new academic positions in rapidly expanding American universities.

Behaviorism

- In Wundt's wake developed the most powerful movement in 20th-century psychology: behaviorism. With behaviorism, the shift is clearly away from both Freud and James and toward the experimental psychology of Wundt.

- Behaviorism carries Wundt's insistence on a science fundamentally tied to experiment and objective observation. Wundt tried to track consciousness and envisaged the behavior of his subjects as a report of introspective experience. Behaviorism went much further, restricting study to only what is observable under controlled conditions.

- If science demands that we stick to what is strictly observable in the lab, and only behavior is strictly observable, the scientific psychology we seek will be a science of behavior. With the rise of behaviorism, consciousness loses its central place. With the rise of behaviorism, consciousness seems to lose *any* place.

- Experimental psychology guided by behaviorism ruled most of the 20th century, with John B. Watson and B. F. Skinner as major figures.

- Of course, no movement occurs without a counter-movement. In the 1960s, a cognitive revolution began as a revolt against the reign of behaviorism. Critical to that revolution was Noam Chomsky, who argued that analysis of verbal behavior in terms of the behaviorists' stimulus and response wasn't going to be enough to explain the complex and structured ways that humans use language.

- We will have to go "inside" in order to understand the cognitive factors running behind the observed behavior. In the merging of a number of fields, cognitive psychology came into its own and morphed into what is now known as cognitive science.

- As the 20th century began to shift to the 21st, still another revolution occurred: the neuroscience revolution. With the explosion of new techniques for brain imaging and brain intervention, cognitive science has joined forces with biology, medicine, and genetics.

Suggested Reading

Freud, "Project for a Scientific Psychoanalysis."

James, "The Stream of Consciousness."

Popper, "Science: Conjectures and Refutations."

Questions to Consider

1. Freud says that a memory, a thought, a realization, or a fear can slip from consciousness to unconsciousness, but still continue to guide your behavior. Give an example from your own experience.

2. James emphasizes the continuity of consciousness—the stream of consciousness. Contemporary critics claim that consciousness is actually quite fragmented, with gaps in time, attention, and content. Which account do you think is right? Could they both be right?

Lecture 13

The Enigma of Free Will

A sense of being able to choose different courses of action is a clear characteristic of our subjective experience. But is it a sense of a freedom that is itself real? Or is our sense of free will merely an illusion? This lecture will focus on both the classic philosophical problem of free will and the way in which questions of free will arise in contemporary scientific research. There is much that the philosophical and scientific approaches have in common. But there are also important ways in which they differ.

The Problem

- The universe is governed by cause and effect. What happens at noon Monday is determined by everything that happened before noon Monday. If what happens at any step is entirely a result of what happens at the step before, everything at every step is determined by what happened before. All history is determined. That's the determinism side of the so-called problem of free will and determinism.

- The other side of the problem is the free will part. Our lives—all lives—involve a series of choices. If you trace back the course of your life, you can map it out like a branching tree diagram of choices faced and decisions made, followed by further choices faced and decisions made.

- The problem is that the two pictures we've painted don't fit together. We think of our lives using the second picture: free will. But the way we think of the universe seems to have the first picture built in: determinism.

- One option when faced with this dilemma is to buy the deterministic picture and to kiss free will goodbye. The universe operates by physical laws written in terms of natural forces and fundamental particles. Free human choice isn't part of the picture. It is at best an illusion.

- Another option is to clutch onto free will, denying the deterministic picture in order to maintain the picture we've painted of free decisions, free choices, genuine responsibility, and sometimes grounded regret.

Scientific Confirmation?

- One of the best-confirmed scientific theories of all time is quantum mechanics, the physics of the very small. We have mathematical formulations of the theory that give us, solidly and reliably, the right quantitative results—often strange and unexpected.

- Things become very confusing when we ask not merely how the formulas operate and what they predict but what they are telling us about how the universe works at a fundamental level. For instance, quantum mechanics doesn't merely tell us that we don't know why a particular atom decays at a particular time. It tells us there is nothing to know. There is no reason why a particular atom decays at a particular time.

- The implication of this scientific picture is that the universe isn't deterministic. Not every event is determined by earlier events: The decay of a particular uranium nucleus isn't so determined. While the 19th-century physics of Newton outlined a deterministic universe, 20th-century quantum mechanics tells us that picture is wrong.

- So perhaps that settles it: Our best science tells us that the universe isn't deterministic after all. There is space for free choice, responsible decisions, and free will after all.

- Some thinkers have made precisely that argument, claiming that quantum randomness allows for free will. But that is a step too far and too fast: If some events happen for no reason, how precisely does that give you free will?

Ancient Debates

- Much of the debate over free will, determinism, and randomness played out long before quantum mechanics arose. Stoicism and Epicureanism were two schools of Greek philosophy that continued into the Hellenistic or Roman period. Although Stoics and Epicureans were both materialists, the two schools had a very different take on free will.

- The Epicureans took something like the quantum line. The Latin poet and philosopher Lucretius gives one of the most complete outlines of Epicureanism we have. Everything is material, including the mind, but the atoms of which things are made occasionally "swerve." Lucretius tempers his determinism with randomness in order to carve space for free will.

- The Stoics, on the other hand, were strict determinists. In full acceptance of an unbreakable chain of cause and effect, everyone's fate is sealed. The best one could do, the Stoics said, was to bear it stoically. That has problems too: If everything is determined, it will be determined whether you suffer your fate or freely accept it. Even there you wouldn't be free.

Roman philosopher Seneca

Compatibilism

- An alternative approach is to reject the dilemma itself: Maybe, once we really understand the philosophical issues involved, we will see that free will and determinism are compatible. Not surprisingly, this is known as a compatibilist approach.

- The basic idea is well captured in a quote from the American philosopher John Dewey: "What men have esteemed and fought for in the name of liberty is varied and complex—but certainly it has never been a metaphysical freedom of will."

- Dewey is emphasizing that metaphysical freedom—freedom from chains of causality—is not the freedom that we care about. What we care about is freedom from tyranny, from oppression, from chains of iron. It may also be freedom from addiction and compulsion. To want those kinds of freedom isn't to want to be metaphysically independent of cause and effect. Freedom in the sense worth caring about is freedom from coercion, not freedom from causality.

Grey Walter's Experiments

- Now let's shift to questions of free will in a scientific context. We will examine two very suggestive sets of experiments.

- The first set of experiments, conducted by the neurologist Grey Walter, date to the 1960s. Walter implanted electrodes in the motor cortex of brain surgery patients. Those enabled him to record a pattern of activity called readiness potential: a burst of electrical activity in the motor cortex that precedes actions like moving your arm, or hand, or finger.

- Grey Walter's hypothesis was that those bursts of activity in the brain didn't merely precede voluntary action. They were the initiation of the causal chain of voluntary action, from brain to hand. In order to test the hypothesis, electrodes were implanted in the area of the brain associated with finger movement.

He then asked his patients to control the movement of an old-fashioned carousel slide projector: When a patient wanted to see the next slide, they'd press the button.

- Grey Walter rigged the slide projector so that what actually triggered the slide change wasn't the press of the patient's finger, but the readiness potential in the motor cortex. There was a direct brain-to-slide projector connection instead of the normal finger-to-slide projector connection.

- The decision to change the slide was still up to the patient. The patient would then reach to press the button in order to change the slide.

- Curiously, that wasn't how Grey Walter's patients experienced it at all. Instead, they reported that the slide projector clicked to the next slide just before they decided to move it.

- The results seem to mean that we had the timing of events wrong. This order is incorrect: First, we have that moment of conscious decision; that activates the readiness potential; and that produces the movement. Instead, the order seems to be: First, the readiness potential starts the chain of events; then we have that moment of conscious decision; and then the finger moves.

Benjamin Libet's Experiments

- A later set of experiments by Benjamin Libet seem to show much the same effect. Libet, like Grey Walter, wanted to know how spontaneous voluntary action worked and to know about the timing of events in the brain and the moment of conscious decision.

- Libet's method was to ask subjects about when they decided to move not at the time of the decision, but later. He asked his subjects to move their right hand any time they felt like it, but to watch a spot of light revolving in a circle as they did so, something like a rotating hand on a clock. Subjects were to remember where the spot of light was when they decided to move their hands. The light served as a timekeeping measure: Each person could report where the light was when they decided.

- The experiment also recorded the timing of the readiness potential in their brains and the moment at which their hands actually move. The findings: The subjective intention or decision to move comes about 200 milliseconds—a fifth of a second—before the movement itself.

- But the readiness potential comes about 535 milliseconds before the movement itself. That's over half a second before. That means the readiness potential that makes your finger move comes before you are aware of a decision or intention to move.

- The simplest interpretation—the one Libet offers—is that your experienced decision or intention can't be what actually makes your finger move. Your brain is well on its way to making your finger move almost 350 milliseconds before you think you are deciding to make it move. Your subjective experience of deciding to move your finger comes more than a third of a second after the process is already in play.

Interpreting

- Do those experiments show that we don't have free will? When we ask that, we're not asking merely what happened in the experiments. We're asking how to interpret what they really mean.

- Libet himself seems uncomfortable with the conclusion that free will is an illusion. Although the readiness potential starts the chain to movement, Libet claims that some of his data show that it is possible for that chain to be consciously interrupted before the movement actually happens. Libet speaks of this as a veto function.

- There are other examples in which consciousness seems a late arriver. Sports excellence often relies on split-second timing. At times, those split seconds are too fast for consciousness to play the kind of role we might expect.

- For a time, the reigning champion in the 100-meter was Linford Christie. In the 1996 Olympics, Christie was disqualified for twice jumping the gun. Video replay clearly shows that the starter pistol was fired before Christie

started to move. Nevertheless, the officials stood by their decision. A similar outcome befell John Drummond in the 100-meter quarterfinal of the 2003 World Championships.

- How is that fair? Both Christie and Drummond moved only after the gun had fired. The answer is that since the 1970s, false starts have been determined with great precision, both electronically and with careful attention to reaction times. Typical reaction time for top-level competitors is something like 125 to 250 milliseconds. If a runner reacts significantly faster than that, they can't have heard or seen the signal and then started. Drummond, for example, moved in less than 100 milliseconds after the signal. He jumped the gun, said the officials.

The Takeaway

- The Walter and Libet experiments do seem to show that a particular picture that we have of free will is often wrong. It's not true that a moment of decision precedes voluntary action in all cases.

- That doesn't mean, however, that the picture is always wrong: A person can decide to marry a spouse long before they pop the question. It is in terms of milliseconds and instantaneous spontaneous action that the picture fails, not in the minutes, hours, days and years of our deliberate action.

- Maybe the standard picture—deliberation first, decision initiating action—is right for the kind of free will we really care about: freedom in major choices, important decisions, even decisions about whether to do the shopping now or later.

- If our picture of free will is one in which every movement in every context has to be initiated by a moment of conscious decision in order to be free, we'll have to conclude that in that sense many of the simple movements we make are not free. But that may not be the kind of freedom we really care about, either.

Suggested Reading

Dennett and Kinsbourne, "Time and the Observer."

James, "The Dilemma of Determinism."

Libet, "Unconscious Cerebral Initiative and the Role of Conscious Will in Voluntary Action."

Questions to Consider

1. Do you think you have free will? In your answer, specify precisely what you mean by free will.

2. If you said yes in answer to question 1, what would convince you that you didn't in fact have free will?

Lecture 14

Emotions: Where Mind and Body Meet

Are emotions aspects of mind that produce a bodily reaction? Or are they bodily phenomena that produce a mental reaction? The answer turns out to be more complicated than either of those options. As a phenomenon where mind and body meet, emotions allow us to trace some of the complex details of mind-body interaction. That's what we'll examine in this lecture.

Defining Emotion

- What precisely do we mean by emotion? Or, what precisely do we mean by emotions? Those two ways of phrasing the central question reflect a very basic division of opinion in the field.

- Many researchers think that we have multiple emotions: distinct modular systems of emotional processing, perhaps reflecting distinct patterns of processing in the brain.
    - They propose a small set of six primary emotions: happiness, sadness, fear, anger, surprise, and disgust. These primary emotions are thought to be universal across human experience: A frown is taken as a universal sign of sadness. Raised eyebrows indicate surprise.

    - An alternative proposal is that there are not six but eight primary emotions. This approach adds acceptance and expectancy to the other six. Whatever the number, more complex emotions, such as pride or disappointment, are seen as a mixture of the primaries.

- On the other side of the fence are researchers who think of emotion in terms of a single multi-dimensional psychological space. Emotions are points of a single psychological space. It is the space as a whole that constitutes emotion.

- Wilhelm Wundt proposed that all emotion could be mapped in three dimensions. One dimension was how pleasurable or unpleasurable the emotion was. One was how arousing or subduing. The third dimension was strain or relaxation. In the 1980s, Havlena and Holbrook proposed a three-dimensional field model for emotion in terms of pleasure, arousal, and dominance.

- Whether framed in terms of multiple primary emotions or in terms of dimensions in an emotional field, most theorists recognize that our emotions are often a complex mixture of simpler elements. We feel guilty pleasure, with elements of both pleasure and guilt. We feel affection streaked with minor irritation, or annoyance tempered by pity.

- Just how universal are the primary emotions that we have outlined? Some critics have claimed that even the basic emotions require a cultural context.

- Once we go from primary to complex emotions, it proves still harder to argue for universality. The German language recognizes an emotion for which English doesn't have a word: *schadenfreude*. We need a whole phrase to express schadenfreude in English: It's "delight at someone else's misfortune."

The James-Lange Theory

- As in so many areas, it is William James's observations and speculations that set the stage for later work on emotion. James thought emotions are not an aspect of mind that then produces a bodily response. Rather, it's precisely the other way around: We don't run because we're afraid; we're afraid because we run.

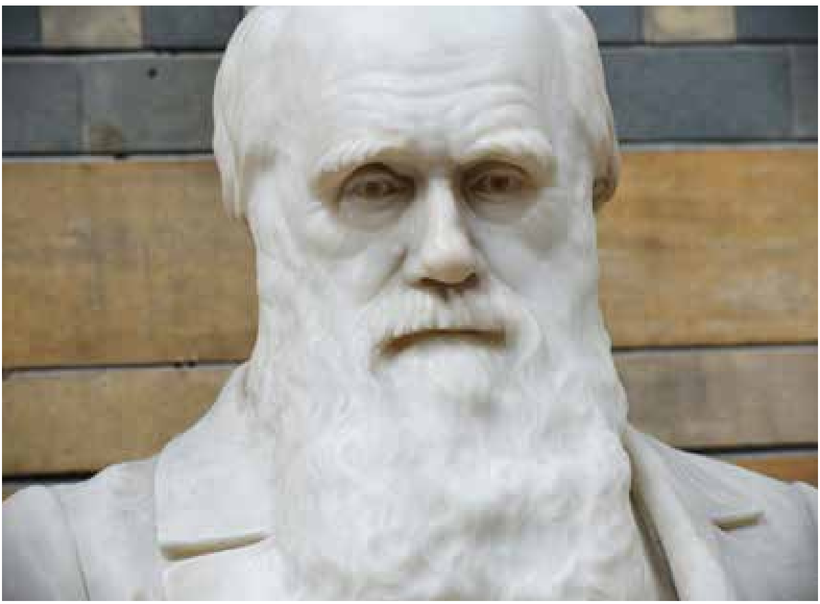
- Here's the way James puts it in 1890: "Common sense says that we lose our fortune, are sorry and weep; we meet a bear, are frightened and run; we are

insulted by a rival, are angry and strike. … [My theory is that the] sequence is incorrect … we feel sorry because we cry, angry because we strike, afraid because we tremble."

- This is called the James-Lange theory of emotion. James and the Danish physiologist Carl Lange didn't work together, but they both developed a body-based theory of emotion at about the same time. In all his writings, James is scrupulous to give Lange equal billing.

- According to James, we can't produce emotion as a purely mental phenomenon. If you try to produce emotions without the initiating physiological cause, you'll get something hollow.

- Here is how James puts it: "Can one fancy the state of rage and picture … no flushing of the face, no dilation of the nostrils, no clenching of the teeth, no impulse to vigorous action, but in their stead limp muscles, calm breathing, and a placid face?" According to James, the answer is no.



British scientist Charles Darwin

- This argument is convincing, but doesn't absolutely set in stone the idea physical reactions have to come before emotions. In order to be fear, the trembling may have to be there. But that doesn't really mean it has to be the instigating cause.

- At the same time, there is some evidence of a link. Moebius syndrome is a rare genetic disorder in which people are born with facial paralysis and unable to move their eyes. They can't frown, smile, or raise their eyebrows in surprise. Not only do those with Moebius syndrome have trouble conveying emotion to others, it appears they have trouble feeling emotion as well.

- The idea that facial expression influences mental state has led to proposals for the treatment of depression. There are clear facial signs of chronic depression in the form of permanent frown lines and a furrowed brow. If we could change those, might we be able to affect the depression itself?

- The answer seems to be yes, with a mounting chain of evidence. Plastic surgeons have reported that individuals who had Botox treatment for frown lines seemed to lighten their moods.

- And a 2012 study conducted in Europe found lower signs of depression in those who received Botox injections over a placebo. Botox works by blocking the neurotransmitters sent to the facial muscles. There can be no frown and so no feedback message of frowning from face to the emotional brain. As James said, "Refuse to express a passion, and it dies."

Criticism

- The James-Lange theory set the stage for later work on emotion. It has also served as a lightning rod for criticism. A major challenge came with the work of the Harvard physiologist Walter Cannon and graduate student Philip Bard in the 1920s.

- Cannon and Bard's first argument against the James-Lange theory was that the same physiological changes are present in very different emotional states.

Increased adrenalin, heart rate, levels of blood sugar, sweating, and pupil dilation occur across many forms of excitement, for example, in both fear and rage.

- Here's their second argument: Artificially producing the physiological states characteristic of specific emotions does not necessarily produce those emotions. Running on a treadmill causes raised heart rate, sweating, and the like. If the James-Lange theory were correct, shouldn't running on a treadmill make us afraid?

- Cannon and Bard surgically cut the connection from the sympathetic nervous system to the brain in cats. In all parts of their bodies that could still express it, these cats continued to show signs of rage when exposed to a barking dog. They concluded that emotion is not simply a response to bodily input. It remains even without that input.

- As an alternative to the James-Lange theory, Cannon and Bard proposed a branching theory of emotion. They proposed that sensory input goes to the thalamus, located essentially at the center of the brain. From there it is sent in two directions. One route goes directly to the body, stimulating the immediate physiological reactions that James noted. The other route goes to the cortex, resulting in the feeling of emotion. In the branching theory, those two routes are independent. Although the feeling of emotion and the body response come from the same place, neither directly produces the other.

- Cannon and Bard didn't succeed in refuting the James-Lange theory on all counts. Cannon and Bard argued that if emotion were physiologically based we should have difficulty distinguishing the two. In fact, it turns out that we do: Sometimes the same physiological input is interpretable in different ways.

Dutton and Aron

- In a classic experiment by Donald Dutton and Arthur Aron, male participants were asked to walk across either a stable bridge or a frightening suspension bridge. At the end of each bridge stood an attractive female experimenter, who

handed the participants an ambiguous picture about which they were asked to write a story. She also gave them her phone number in case they had any further questions about the experiment.

- The men crossing the scary bridge made up stories that had a higher sexual content than those by the men crossing the stable bridge, and the scary-bridge crossers made more follow-up phone calls to the attractive experimenter.
- Emotion doesn't appear to be a merely physiological stimulus. It doesn't appear to be merely a branching pattern to independent physiology and feeling as suggested by Cannon and Bard. Emotion appears to have a contextual and interpretational element, involving both physiology and cognition.
- That result supports Stanley Schachter and Jerome Singer's two-factor theory of the 1960s. Schachter and Singer claimed that our emotional response often incorporates two elements. One part is physiological arousal. Another part is the cognitive label that we assign to that arousal on the basis of context.

The Full Picture

- What has now become the standard picture of emotional processing incorporates bits and pieces from the entire history we've traced. The neuroscientist Joseph LeDoux has been a major figure in the development of that standard theory.
- At the core is the branching hypothesis of the Cannon-Bard model. The stimulus of the charging bear is directed to the sensory thalamus. From there it goes in two different directions: to the amygdala, identified as the emotional instigator, and to the prefrontal cortex, where emotion is felt for the first time.
- The prefrontal cortex is identified with deliberate judgment and decision, and this is where aspects of the Schachter-Singer model come in. It is the prefrontal cortex that interprets the emotion, shaping and perhaps inhibiting response.

- From the prefrontal cortex, signals are sent back to the amygdala. Those may be reinforcing: "Run faster." They may be corrective: "Wait, I can't outrun a bear. Should I freeze?"

- Regarding the James-Lange theory, the neuroscientist Antonio Damasio takes pains to defend that model against the Cannon-Bard arguments. James's theory, he says, was unfairly attacked and dismissed.

- Although it's not true that the body is the only source for feelings, James was right that feelings can be a reflection of body-state changes. As an example, patients with damage to the spinal cord often do report a damping down of emotion. The higher the damage to the spinal cord, the more seems to be lost.

Emotion and Reason

- From the Greeks, we inherit the view that emotion is a threat to rationality. Plato characterized reason as a charioteer, needed in order to keep the wild horses of emotion under control.

- Aristotle defines man as a rational animal, with the rational principle in charge of any lower impulses. The Stoics tell us that the best life is one in which we rationally remove ourselves from the turmoil of the passions.

- In the contemporary picture, it is in the cortex rather than in the amygdala that the feeling of emotion occurs. Damasio reports on a patient called Elliot in whom that route to the cortex was lost.

- The history of philosophy would suggest Elliot's case to be the philosophical ideal: a life of reason freed from disruptive emotion. That's not the case: Without a route from the inner brain to the cortex—without felt emotion—Elliot was lost. He performed normally on all intellectual tests, on tests of background knowledge, and on tests of reasoning, even ethical reasoning. But Elliot could no longer make appropriate decisions.

- Without being able to register emotions, Elliot wasn't able to evaluate alternative courses of action; he couldn't make up his mind regarding personal and social matters. He lost his job, one marriage, and then a second marriage, along with his sense of responsibility.

- There is now a wide consensus that rationality requires an emotional component. In 1983, Howard Gardner introduced a theory of multiple intelligences, attempting to expand a concept of rationality beyond the confines of IQ. EQ, or emotional quotient, has been proposed as a necessary supplement.

Suggested Reading

Damasio, *Descartes' Error*.

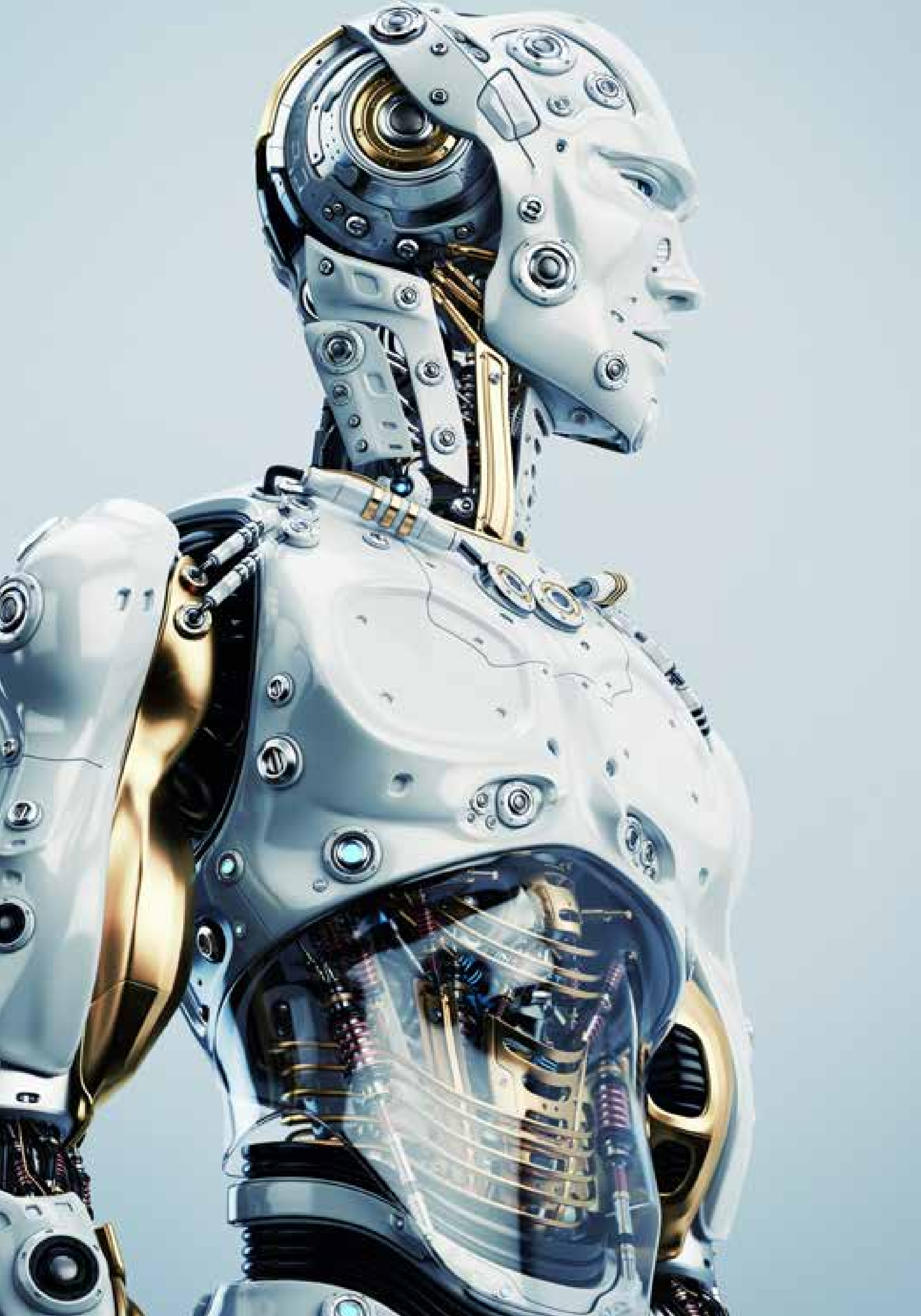
James, "What Is an Emotion?"

Thagard, *Hot Thought*.

Questions to Consider

1. Make a list of at least 25 emotions. Are some of these compounds of others? Do some, like awe or pride, demand a particular kind of object or context?

2. Darwin says that bodily expression of an emotion amplifies and intensifies it, whereas dampening the bodily expression can soften the felt emotion as well. Give an example from your experience where that does fit, and perhaps an example where it does not.



Lecture 15

Could a Machine Be Conscious?

In 1939, at the University of Cambridge, there were two courses offered on the foundations of mathematics. One was a course on the foundations of mathematics by the influential and charismatic philosopher Ludwig Wittgenstein. The other was a course on the foundations of mathematics by the mathematician Alan M. Turing, who in the previous 10 years had laid down some of the theoretical work that led quite directly to computers. This lecture tracks some of the major ideas of those two figures, Wittgenstein and Turing. Their ideas were central to the fields of 20th-century philosophy of mind, computer science, and artificial intelligence.

Wittgenstein and Turing

- Most philosophers develop only one major viewpoint, in one extended period of exploration. Wittgenstein developed two contrasting viewpoints in two distinct periods. The first viewpoint is that of his *Tractatus Logico-Philosophicus* of 1921, which offers a vision of the logical structure of the world mirrored in the logical structure of language.

- By 1939, Wittgenstein was well into his second period of intellectual exploration. He renounced the vision of the *Tractatus*. In his second period, Wittgenstein offers a very different picture of language and the world, a picture in which it is the linguistic practice of a social community that is the key to understanding. The slogan for this second period is "meaning is use."

- Turing was 23 years younger than Wittgenstein, born in London in 1912 as the second son of a British civil servant with a commission in India. His parents traveled often between England and India, leaving the boys with an older couple or in their private schools.

- In school, Turing's genius for mathematics was discovered and cultivated early on. By 1939, Turing had completed his major theoretical work in what is now regarded as the foundations of computer science.

- In his lifetime, Wittgenstein published only the brief 75 pages that constitute the *Tractatus*, one article, one book review, and a dictionary for children. Despite that thin record of publication, his impact was enormous, based largely on the force of a charismatic personality and a dramatic air of ascetic philosophical intensity. By contrast, Turing was a stutterer, careless in dress and cleanliness, and entirely undiplomatic in expressing his opinions.

- Their ideas on the foundations of mathematics were as diametrically opposed as their personalities. Turing thought of mathematics as something that was essentially discovered, something like a science of the abstract. Wittgenstein insisted that mathematics was essentially something invented, following out a set of rules we have chosen—more like an art than a science.

The Language Argument

- About the time of the Cambridge lectures, Wittgenstein developed a central component in his later thinking: the private language argument. One way to read it is as an argument that there can be no private language. Our language is a public language, with a public use: learned, applied, and corrected in a social context of communication.

- Wittgenstein uses the parable of the beetles in the boxes to illustrate his claim. Imagine two people, each with a beetle in a box. Using their respective private languages, one person might describe their beetle as "kufu kufu" and the other might describe their beetle as "niki niki." If they can't see each other's beetles, those words will never acquire meaning.

- The learning process is entirely public, and the correction required in learning demands that it be public. For instance, a child who identifies a cat as a truck has to be corrected. If our mental terms referred to something inherently private and locked way, we couldn't ever learn those terms.

- Wittgenstein's private language argument is best conceived as a negative argument. It's an argument against private Cartesianism. However, many of his students at the 1939 lectures interpreted it as a positive argument for a particularly extreme form of behaviorism.

- Behaviorism dictates that data is first and foremost behavioral data: Any science of the mind must at base be a science of behavior. Wittgenstein's students in those 1939 sessions took the private language argument not merely as a negative argument against Cartesian privacy but as a positive argument for analytic behaviorism. Their view was that talk of the mind must ultimately simply be talk of behavior.

- Wittgenstein emphasized that our mental terminology is part of our way of dealing with each other. It has the meaning it does as part of language use in a community. Because of that tie to a community of creatures like us, Wittgenstein seemed to think that it's a category mistake to speak of machines as thinking, let alone to think of machines as conscious.

Turing on Machines

- In 1939, Wittgenstein was emphasizing the role of language—including the role of mental terms—as part of an essentially human way of life. By 1939, Turing had developed something very different: a fascinating and fully developed theoretical picture of machines and their prospects for computation.

- Turing was a mathematician. He wanted to know what aspects of mathematics were computable and what weren't: Are there limits to what is algorithmically computable?

- He wanted an answer to that question about mathematics that was as tight as mathematics itself. In order to get that, he needed to have a formal

definition of these intuitive things: calculations, computations, algorithms, and effective procedures. Turing realized that by formalizing the notion of a symbol-manipulating machine he could get a rigorous handle on questions of computability and its limits.

- The abstract model he developed is what we know as a Turing machine, for which he was already well known in those lecture sessions in 1939. It is still regarded as the standard model for computability in general.

- A Turing machine is not a physical thing. It is an abstract model of an information-processing machine. It has some number of possible internal states. It can be in state 1, or state 1,207, but it has some certain number of possible states it can be in at any given time.

- Its information input is a set of symbols written one by one in squares on a tape. We won't give it memory limitations—the tape extends as far as we need in both directions.

- Here's how the machine works: At any point in time it sits on just one square of the tape and operates in terms of just two things: the symbol beneath it on the tape and the internal state it happens to be in.

- What it does when it sees a particular symbol and is in a particular state is dictated by its rules; for example, if it sees the symbol *a* and is in state 2,007, it will do a certain thing, and if it sees the symbol *c* and is in state 45, it will do a different certain thing.

- Today it's hard to believe how revolutionary Turing's model was. After all, it sounds like "just" a computer. The point is there were no computers when Turing developed his abstract model. The real machines came later, based on Turing's abstract vision of information processing.

- Turing envisaged information processing as symbol manipulation. His model was conceived purely theoretically but ended up with very practical consequences. All of our contemporary computers trace back to Turing's vision.

- In World War II, Turing put his theory to practical use, developing computing

machinery that managed to crack the Nazi Enigma code. After World War II, Turing was instrumental in the construction of larger and all-purpose computing machines. But his vision went even farther. In his notes, Turing anticipated the next generation of computing: neural networks, inspired by the way that neurons function in the brain. Neural networks show an amazing ability to learn.

- The end of Turing's story is not a happy one. Security in the development of computers was tight after World War II. The Americans were wary of two groups they thought of as security risks: communists and homosexuals. Turing was homosexual. He was forthright and honest about his inclinations as he was about everything else. In England at the time, homosexual conduct was a crime. In 1952 he was arrested, tried, and convicted.

- It was clear that his security clearance was in danger—that he would be shut out of working in the field that he had essentially invented. In 1954, he committed suicide by eating an apple dipped in cyanide.

Turing on the Mind

- The field of artificial intelligence is another part of Turing's legacy. It is here that we come back to his views on the mind. In 1950, Turing published his definitive philosophical piece on the question of minds, machines, and information: "Can a Machine Think?"

- Turing rejects the question of if a machine can think as meaningless but proposes replacing that question with one that isn't meaningless. It's called the Turing test: Could you build a machine that under specified conditions was indistinguishable from a person?

- Note that the Turing test is an entirely behavioral test. It is the behavior of a properly programmed machine that Turing predicts will be indistinguishable from that of a human.

- Turing's article and the Turing test were a major instigation for the field of artificial intelligence, a term that wasn't invented until after Turing's death in 1954.

- In 1956 a group of collaborators put together the Dartmouth Summer Research Project on Artificial Intelligence. The theme was clearly along Turing's lines of inquiry. The conference was explicitly based on the conjecture "that every aspect of learning or any other feature of intelligence can in principle be so precisely described that a machine can be made to simulate it."

- There were actually two branches of artificial intelligence that came out of the conference. One involved straight programming, on the model of the Turing machine. The other involved the first explorations in neural nets, another line of thought that Turing had anticipated.

The Legacies

- In 1950, Turing predicted that by the year 2000 we would have machines that passed the Turing test: machines good enough at simulating human intelligence that an average interrogator wouldn't have better than a 70 percent chance of distinguishing mind from machine after a five-minute exchange. Turing was wrong about that. We're long past the year 2000 and still don't have computers that can pass the Turing test.

- But there was another thing Turing was certainly right about. He said that he expected language to change: "I believe that by the end of the century the use of words and general educated opinion will have altered so much that one will be able to speak of machines thinking without expecting to be contradicted."

- He was absolutely right about that. Today we have no problem using a full range of anthropomorphic concepts in order to characterize what our computers do. We say "It can't see items in that subdirectory," and "It thinks I put that document somewhere else." Turing was right: We speak of our machines as knowing things, thinking things, even forgetting things or ignoring us.

- Wittgenstein didn't see that far ahead. Today it looks like Wittgenstein's refusal to speak of machines thinking, at least in today's informal and colloquial sense, represented an outlook appropriate to Cambridge in 1939.

- Then again, maybe Wittgenstein was right in saying that terms like *thinking* are appropriate only for particular kinds of agents in a particular way of life. What he wasn't able to see was that our way of life would change in ways that would give machines a much more important and integrated role.

Suggested Reading

Casti, T*he Cambridge Quintet*.

Diamond, ed., *Wittgenstein's Lectures on the Foundations of Mathematics Cambridge 1939*.

Turing, "Computing Machinery and Intelligence."

Wittgenstein, *The Blue and Brown Books*.

Questions to Consider

1 Which of these do you think is possible for a machine? In each case, specify why or why not:

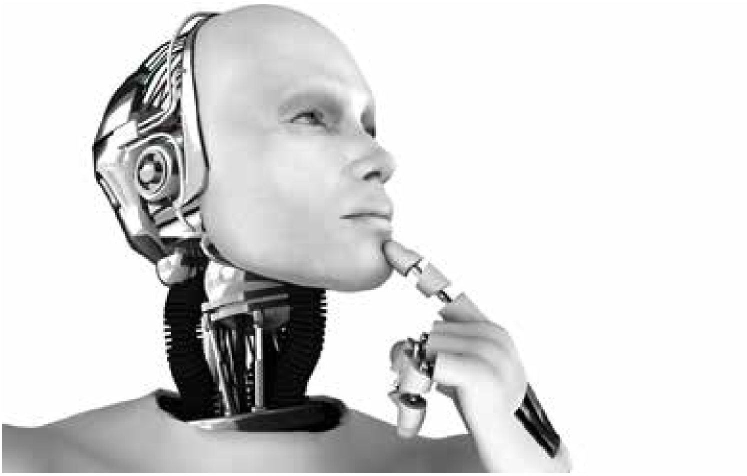
- ☐ calculate
- ☐ compute
- ☐ respond
- ☐ perceive
- ☐ remember
- ☐ think
- ☐ reflect
- ☐ understand
- ☐ be conscious

2 Do you think it will be possible to build machines that pass the Turing test? If so, by what date would you guess?

Lecture 16

Computational Approaches to the Mind

This lecture explores two computational approaches to the mind. One is a computational approach to the mind through artificial intelligence. The other is a computational approach through the concept of information. The first approach asks: Can we understand intelligence better by attempting to build intelligent machines? The second approach asks: Can computational approaches take us as far as really understanding intelligence?



The Beginning

- A key figure in the history of both approaches is Alan M. Turing, who died in 1954 by eating an apple dipped in cyanide. Two years later, the Dartmouth conference was held. It was the first devoted to what they decided to call artificial intelligence.

- There were actually two groups of people involved in early artificial intelligence. One group wanted to pursue artificial intelligence in order to understand intelligence. Intelligent machines were a means to an end.

- The other group simply wanted to build smart machines. If understanding human intelligence would help in that task, fine. But understanding human intelligence was simply a means to that end.

- As well as two different goals, there were two different methodologies. Some of the participants at the Dartmouth conference intended to produce intelligence using straightforward linear computer programming. The other methodology was to copy the mechanisms of the brain.

- For many years, it looked like the route to modeling intelligence was straightforward computer programming—known today colloquially as "good old-fashioned artificial intelligence," or GOFAI.

GOFAI

- Allen Newell and Herbert A. Simon of Carnegie Mellon each had a background that combined psychology with computer science. At the Dartmouth conference, they showed off their brand new Logic Theorist program.

- Simon and Newell's program would eventually prove 38 of the first 52 theorems of Bertrand Russell and Alfred North Whitehead's *Principia Mathematica*, a landmark in 20th-century logic.

- The next year Newell and Simon unveiled the General Problem Solver, designed for a wider class of problems. Herbert Simon claimed that they had "solved the venerable mind-body problem, explaining how a system composed of matter can have the properties of mind."

- With that first flush of success came bold predictions for the future. In 1958, Newell and Simon predicted that "within ten years a digital computer will discover and prove an important new mathematical theorem." And "within ten years a digital computer will be the world's chess champion." Neither happened. In 1965, Simon predicted that "machines will be capable, within 20 years, of doing any work a man can do." That didn't happen either.

- By 1974, the over-hype caught up with the field. The programming approach hit a wall. It's known as Moravec's paradox, after the researcher Hans Moravec. The initial tasks which artificial intelligence attempted—

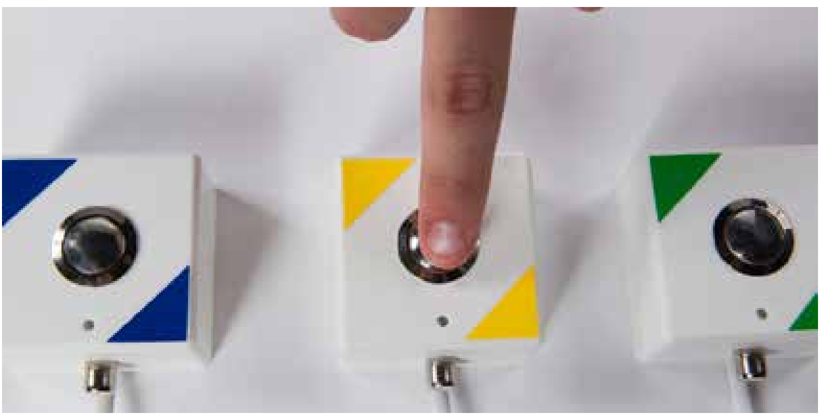
proving theorems and solving geometry problems—turned out to be the easy problems for machine programming. By contrast, a range of tasks that are easy for us—pattern recognition, facial recognition, maneuvering across a crowded room—turned out to be forbiddingly hard to program into machines.

The Second Approach

- What of the second methodological approach to artificial intelligence—using artificial neural nets in order to build intelligent machines on the model of the brain?

- In the late 1950s, the psychologist Frank Rosenblatt developed a computer network that functioned roughly like a network of neurons. A node in the network receives input from various other nodes; when it reaches a certain threshold, it fires; then that firing sends input to other nodes down the line, just like a neuron does.

- The real breakthrough was the fact that Rosenblatt's neural networks could learn. By repeating what answer was wanted from a set of sample inputs, a neural network could make its own adjustments to produce the desired result. Rosenblatt called his neural networks perceptrons.



A contestant asked to choose between two prizes is facing an "exclusive or" choice.

- Perceptrons consisted of just two layers of nodes: a set of input nodes connected to a set of output nodes. The training scheme that Rosenblatt used demanded that simplicity. It only worked for two-layer networks.

- Precisely because of that simplicity, it turned out there was also something that perceptrons couldn't learn. It was impossible for a perceptron to learn what is called an "exclusive or": This is true, or that is true, but not both.

- The fact that perceptrons had that limitation was the key point in a devastating critique of neural networks leveled by Marvin Minsky and Seymour Papert in their book *Perceptrons*. Minsky was a proponent of the GOFAI approach.

- Minsky and Papert's critique resulted in research funding and efforts being directed almost exclusively to the GOFAI approach. It set back by decades the attempt to build smart machines on the model of the human brain.

- Frank Rosenblatt died shortly after the book was published, on his 43rd birthday, in what was reported as a boating accident on the Chesapeake Bay.

- In the long run, neural networks came back strong. In the 1980s, the physicist John Hopfield introduced a form of neural nets that could learn in an entirely new way. The psychologists David Rumelhart and James McClelland also developed a new form of neural net learning that could be applied to more than two layers. Neither "exclusive or" nor any other function of inputs were a problem any longer.

Advances

- What have we learned in the process? Have machines taught us what intelligence is? Can we build truly intelligent machines?

- In 1997, IBM's Deep Blue became the first computer to beat a reigning world chess champion, Garry Kasparov. That was a good 30 years after Newell and Simon's prediction. In order to beat Kasparov, Deep Blue reportedly scanned 200 million possible moves every second. We know that human chess masters don't think that way.

- In 2011, IBM's computer Watson beat the two greatest *Jeopardy!* champions at their own game. How? Watson had access to 200 million pages of content, including the full text of Wikipedia. It searched word frequencies that would fit a basic syntax of questions and answers.

- Despite being able to generate correct questions quickly, Watson had no idea what it was saying. Here is one clue Watson responded to: "It was the anatomical oddity of U.S. Gymnast George Eyser, who won a gold medal on the parallel bars in 1904." Watson said, "What is a leg?" Watson wasn't awarded credit. The answer is, "What is a missing leg?"

- David Ferrucci, head of the Watson project, said that Watson didn't understand the idea of "anatomical oddity" in the original clue.

Intelligent Machines

- What of the future of intelligent machines? Consider Moore's law, named for Gordon Moore, the cofounder of Intel. Moore noted that the computing power of available hardware doubles about every two years.

- That exponential growth in computing power is the basis for predictions by Hans Moravec and computer scientist and futurist Ray Kurzweil that we are fast approaching a point at which our machines become more intelligent than we are.

- Kurzweil extrapolates Moore's Law to predict a future singularity. That is a point at which technological development will be driven by the machines themselves, with progress so rapid that mere humans will no longer be able to comprehend it. Kurzweil says the singularity will be here by 2045.

- Moravec extrapolates Moore's law in a similar way to predict a "mind fire" of computational intelligence. Our real descendants, he predicts, will not be biological. They will be intelligent machines.

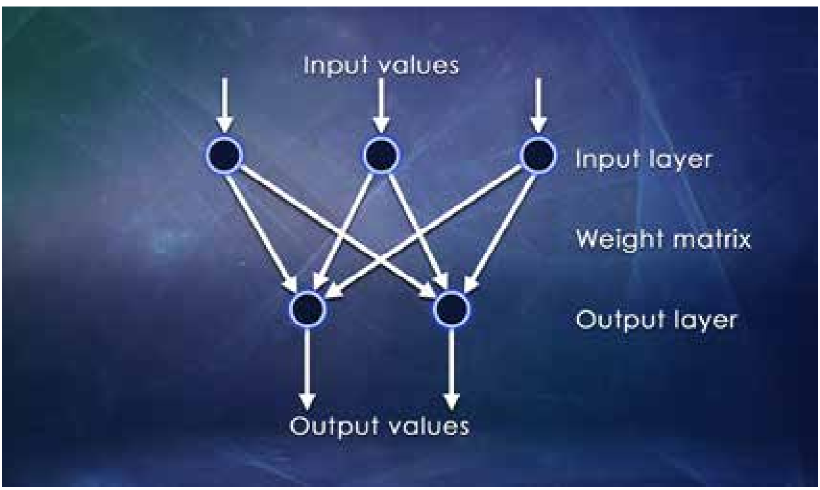
- Both predictions leave unanswered what intelligence is, in either people or machines. Both predictions also rely on a simple extrapolation from a few decades of progress in computer hardware. Moore himself said in 2015, "I see Moore's law dying in the next decade or so."

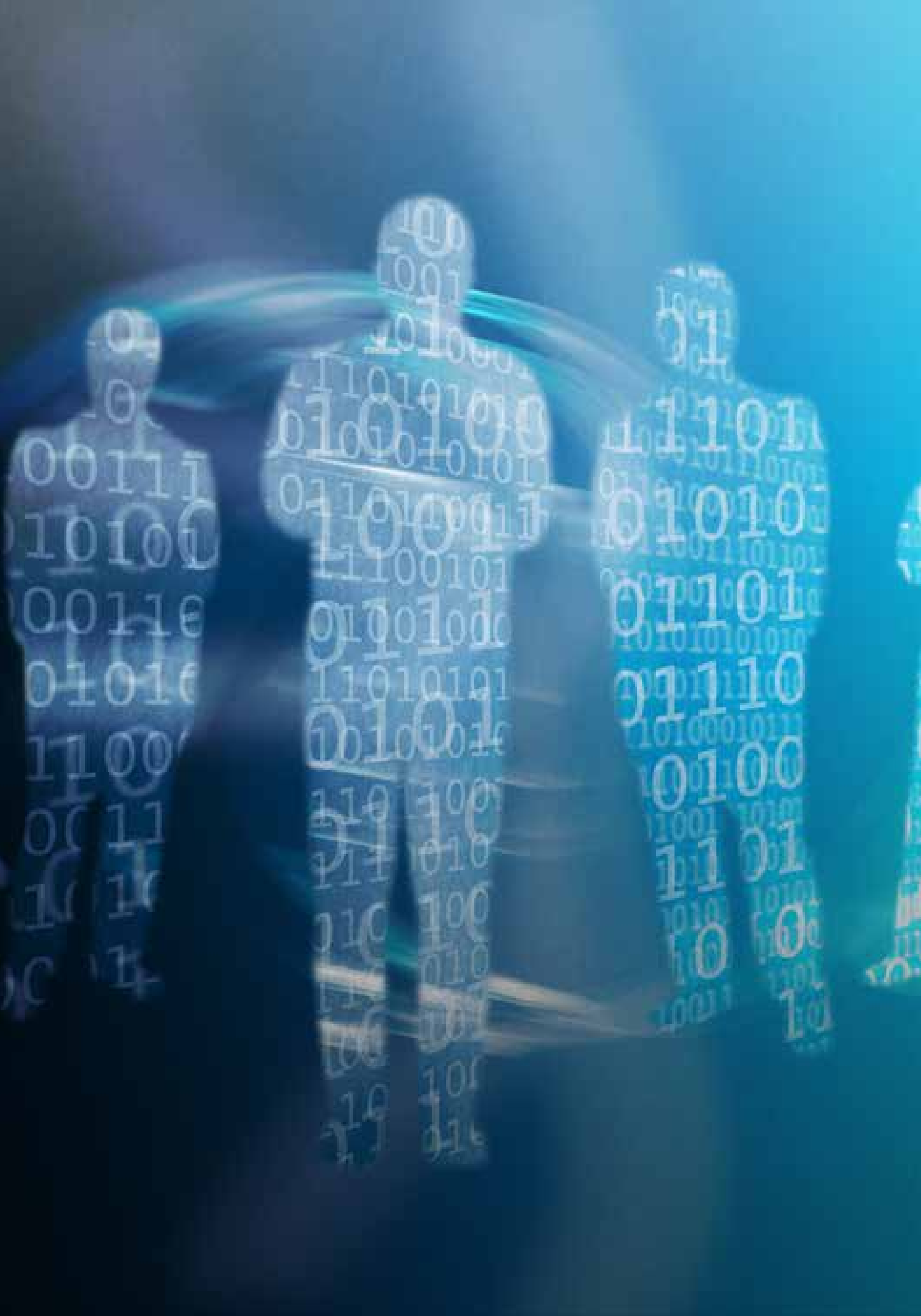
Information

- Consider the brain as an information-processing machine. Can we get a better take on how brains and minds work by understanding information?

- First, what exactly is information? Here again, Alan M. Turing may have played a role. For several months in 1943, Turing was sent to Washington to share the breakthroughs that the British had made in cracking the Nazi Enigma code. He made contact with a young researcher named Claude Shannon at Bell Labs. Turing and Shannon had long conversations over lunch. Information theory may have begun in those conversations.

- The founding document of information theory came a few years later, in Claude Shannon's "Mathematical Theory of Communication." He was still working for Bell Labs. The application he had in mind was how much information you could condense into a signal across a telephone wire.





- At the core of Shannon's information theory is the concept of surprise. The more a message surprises you, the more information it contains. Shannon's measure depends on the number of options that a message eliminates. The more options are eliminated, the greater the information content.

- Consider a coin toss. A coin can only come up one of two ways: heads or tails. If someone can reliably tell you which way it will land, they have eliminated one option out of two.

- Now consider the roll of a die. If a person can reliably predict that, they have eliminated five options out of six. They have given you more information.

- But Shannon's theory encounters a paradox: Under Shannon's rules, a completely random string of 0s and 1s contains the most information as compared to a string of all 1s or 0s, or a patterned string of 1s and 0s.

- Shannon himself recognized the limits of his information theory. At the beginning of his classic paper, he notes that message information often has to be understood a different way: in terms of message meaning. But for his purposes, he just puts that sense of information aside: "These semantic aspects of communication are irrelevant to the engineering problem."

The Global Workspace

- How does information come together in the brain? Bernard Baars is a theoretical neurobiologist at the Neurosciences Institute in La Jolla, California. He has proposed what he calls global workspace theory. It's a theory of how the brain handles information and a theory of consciousness.

- Much of the brain's functioning is unconscious, distributed in different areas. Baars says that consciousness is a "blackboard" or a "global workspace" of working memory that has access to those different areas.

- Consciousness, he says, "resembles a bright spot on the stage of immediate memory, directed by a spotlight of attention under executive guidance. Only the bright spot is conscious."

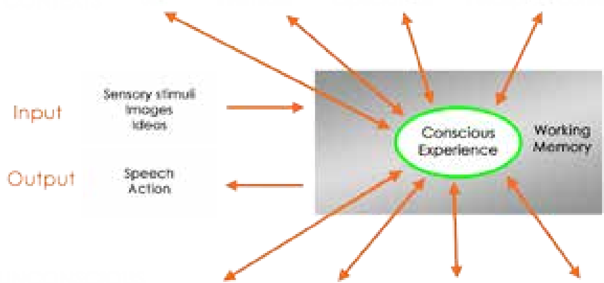
Integrated Information

- There is a recent information-theoretic model of consciousness that builds on both Baars's global workspace and Shannon's information theory. Here the foundational document is a piece by Giulio Tononi at the University of Wisconsin: "Consciousness as Integrated Information: A Provisional Manifesto."

- For Tononi, consciousness is integrated information. Tononi combines Shannon and Baars and offers a mathematical measure for integrated information. He calls that measure phi.

- Consider the sensory chip of a digital camera. It incorporates a million or more photodiodes. That's a lot of information in Shannon's sense. But the information isn't integrated in Baars's sense. It gets a low score using Tononi's phi computation. Furthermore, the sensory chip isn't conscious.

- In contrast, your visual system deals in integrated information. Information comes to you from the millions of photosensitive cells in your retina, but it comes in as an integrated picture of a rushing roadside or a smiling face. It's so integrated that you can't tell what any particular retinal cell is doing. Unlike the camera, your information is integrated information and would score high on phi. And of course, you're conscious.



- However, computer scientist Scott Aaronson points out that there are many simple systems that come out high on Tononi's phi-scale but that no sane person would consider conscious. One of his examples is the operation of error-correcting codes in compact discs.

- These codes work using a two-dimensional grid of very simple logic gates. Despite that simplicity, the system is very highly integrated, so it scores high on a measure of phi. If consciousness were merely integrated information, CDs would be conscious.

Suggested Reading

Kurzweil, *The Singularity Is Near.*

Minsky and Papert, *Perceptrons, Expanded Edition*.

Shannon, "A Mathematical Theory of Communication."

Tononi, "Consciousness as Integrated Information."

Questions to Consider

1. Moravec's paradox is that aspects of what we think of as intellectually challenging—construction of mathematical theorems, for example—turn out to be relatively easy for computers. It is the easy things for us—recognizing faces, for example—that turn out to be hard for computers. Why do you think that is?

2. We talk about information all the time. We are in an information age, after all. But what is information? If asked for a definition, what would you say?

Lecture 17

A Guided Tour of the Brain

Throughout this course so far, we've considered the brain in some detail, but almost always in bits and pieces. To understand brain modularity, we've explored areas in the visual system, examined the fusiform gyrus and its role in prosopagnosia, discussed the role of the hippocampus in memory formation, and talked about the two hemispheres in split brains. This lecture will take on a more holistic and more ambitious project: a tour of the whole brain.

The Brain

- The human brain has been described as the size of a coconut, the shape of a walnut, the color of uncooked liver, and the consistency of chilled butter. It has also been described as the most complex three and a half pounds of matter in the universe.

- Under the microscope, a cross-section of the brain reveals two kinds of cells. The most common are glial cells. *Glial* means "glue." The function of the relatively simple glial cells has long been thought to be primarily structural.

However, like other long-held assumptions regarding the brain, we should be prepared to be wrong. It has also been suggested that glial cells may play a role in amplifying or synchronizing electrical activity within the brain.

- The other cell seen under the microscope is the neuron. The neuron is designed to transfer electrical potential through complicated paths of other neurons. There are about 100 billion neurons in the brain, as many as there are stars in our galaxy. Each neuron has something like 10,000 connections to other neurons.

- The textbook picture of a neuron is of a cell body that receives input from other neurons at a number of dendrites and sends a signal down an axon to the dendrites of other neurons. Neurons don't actually touch: Their contact is across the tiny gap called the synapse. The process of information transfer across a series of neurons goes from electrical to chemical to electrical.

- Just as it combines chemical and electrical processes, the neuron combines analog and digital information. The input gathered by dendrites builds smoothly, essentially in analog form. But when that input crosses a certain threshold, the cell fires, sending what is essentially a digital signal down the axon.

- The picture we've painted so far is really one of excitatory neurons, which stimulate the firing of further neurons at the ends of their axons. But there are also inhibitory neurons, whose work is to impede the firing of those they come in contact with. The ratio is about 80 excitatory to 20 inhibitory.

- Aside from 100 billion neurons, the brain also has synapses in the trillions. Some 50 neurotransmitters are also part of the mix; examples include serotonin, which affects mood and anxiety, and acetylcholine, which is associated with learning, attention, and memory.

- Have we found consciousness yet? No. However necessary neurons are for consciousness, no-one thinks that consciousness exists at the level of a single neuron.

The Brain's Organization and Evolution

- The brain is built on the end of the spinal cord and brain stem, a tube that originally developed in fish in order to more centrally process nerve impulses from various parts of the body. What was originally just a bulge on the top of the spine developed more specialized modules, for instance, modules for smell, sight, and the control of bodily movement—the cerebellum. Taken together, the smell area, primitive visual processing, and the cerebellum constitute the inner core of our brains, known as the reptilian brain.

- Do we find consciousness in the reptilian brain? The cerebellum might be the best candidate for consciousness in that part of your brain. Your body positions itself and moves as smoothly as it does, with balance and with coordinated movements, only because of the cerebellum. But bodily movement isn't initiated in the cerebellum. In your brain, movement is initiated higher in the brain. It is merely coordinated and fine-tuned in the cerebellum.

- The cerebellum's best claim to that kind of consciousness is motor learning. The body memory of how to swing a baseball bat or ride a bike may involve the cerebellum. But once learned, those too are things that we think of as operating automatically or unconsciously. To that extent, at least, your reptilian brain seems an unlikely place to find consciousness as we know it.

- The mammalian brain developed from the reptilian, adding a number of further components. Modules developed as the amygdala (the emotion center), the hippocampus (critical to laying down new memories), the thalamus (where central processing occurs), and the basal ganglia. Together these are known as the limbic system.

- Is it in the mammalian brain that we find consciousness? We are certainly conscious of memories, which the hippocampus lays down. We are certainly conscious of emotions, for which the amygdala is crucial. Signals through the thalamus appear to be part of our attention system and are crucial to various aspects of sensation and motor control of which we are certainly conscious.

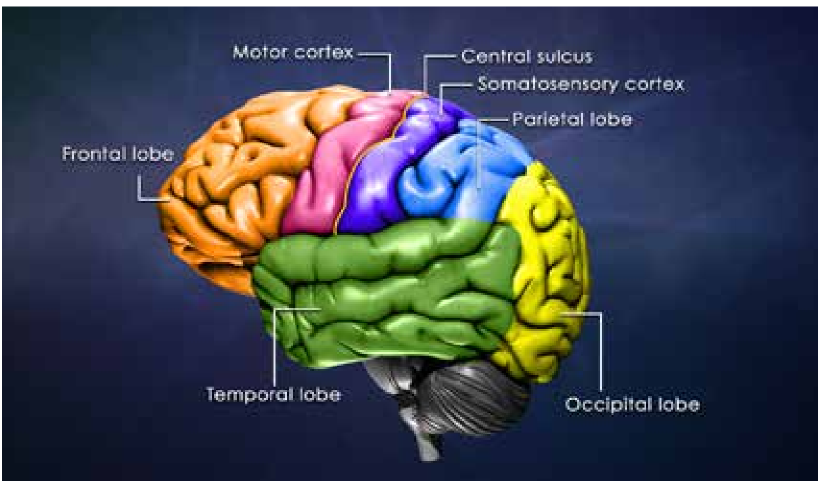
- The mammalian brain is thus clearly important to consciousness as we know it. Like the reptilian brain, however, the mammalian brain is generally thought as operating largely unconsciously. Although we sense our emotions in consciousness, it's rare that we're able to consciously direct them.

The Neocortex

- The mammalian level of our brains is wrapped in a still larger structure: the neocortex. It was originally a skin of cells over the mammalian brain and developed into a massively larger enveloping structure. Roughly three-quarters of your brain is neocortex.

- It is the neocortex that forms the walnut-like exterior of the brain. The wrinkles are described in terms of their canyons—sulci—and ridges—gyri. The neocortex in rats and small mammals, on the other hand, is smooth.

- The neocortex has six layers. It is presumed that the layer structure reflects some form of information processing, but no one knows quite how or why. The neocortex also seems to be organized in terms of columns: particular vertical patches six layers thick. Again, there is a notable lack of consistency in the literature regarding the definition or function of those distinct columns.



- What do the parts of the neocortex do? The occipital lobe, at the back of the head, is largely devoted to the progressive stages of visual processing.

- Sound is the province of the temporal lobe. Although sounds go to both hemispheres, those detected by your right ear are primarily processed in your left brain, while those detected by your left ear are processed in your right brain.

- The processing of speech sounds is concentrated in the left hemisphere. For that reason, a person who goes deaf in the right ear may develop problems hearing language. A person who goes deaf in the left ear may have no problems with language but may lose an appreciation for music.

- Language production and processing is accomplished in two interestingly different areas, both in the left hemisphere but in different lobes of the brain. Wernicke's area lies in the temporal lobe and is where speech is understood. Broca's area is where speech is produced. It lies in front of Wernicke's, across a major canyon called the lateral sulcus.

- Between the two areas, running down and below the bottom of the canyon, is a large bundle of fibers: the language loop between production and processing, Broca's and Wernicke's.

- In the 1860s, a French surgeon named Paul Broca met a man called Tan. Tan was called that because "Tan" was all he seemed able to say. When Tan died, Broca dissected his brain and found a sizeable lesion in what we now call Broca's area. When the Broca's area is impacted, people can understand what is said to them, and know what they want to say, but are unable to say it. The result is often telegraphic speech, individual words that lack the form of grammatical sentences.

- If Wernicke's area is impacted, the Broca's area may still allow a person to produce what sound like full and perfectly grammatical sentences. But because the person cannot understand speech, they can't understand their own.

- Sound processing is a major function of the temporal lobes, but not the only function. The middle temporal gyrus area of the temporal lobes is crucial to judging motion. Face recognition is a function of the fusiform gyrus at the bottom of the temporal lobe. Aspects of memory live here as well.

Development and Damage

- Neural development of the frontal lobes starts at the age of about six months. At that age, the neocortex begins to exert control over the mammalian brain. Before six months, a baby presented with two toys will try to grab both. By the age of one, she will make a choice.

- The ability to understand speech matures before the ability to produce speech. There is therefore a period in which toddlers can understand much more language than they can express. It has been speculated that that is part of the problem with the "terrible twos."

- The frontal lobes are associated with planning, logical reasoning, and problem solving, particularly problem solving that demands creative flexibility rather than standard solutions. Aspects of personality and memory are also credited to the frontal lobes.

- But a core function is management and control of other aspects of the brain—particularly emotion. Here the classic case is Phineas Gage, a railroad foreman in the mid-1800s whose frontal lobes were massively impacted by a mistimed explosion that sent a steel rod through his head.

- Astoundingly, Phineas survived. But he was a changed man. He lost the judgment and executive control that is the function of the frontal lobe. He was converted from a reliable and industrious worker to an aimless drifter.

Suggested Reading

Bear, Connors, and Paradiso, *Neuroscience*.

Carter, *Mapping the Mind*.

Questions to Consider

1 Have you ever wondered whether your brain might function differently than other people's? With the background covered in this lecture, what specific parts of your brain do you think might be different?

2 In the search for how the brain produces consciousness, what do you think we should look for in the evolutionary record? What should we look for in infant and child development?

Lecture 18

Thinking Body and Extended Mind

One could characterize a great deal of the history of philosophy in terms of just two questions. Both are questions of how we can cross a divide. The first question is how we cross the divide between the mind and the world: How can we possibly know what the world is really like, beyond our experience? The second problem is the one we've been tracking throughout this course: How can we understand the relation between our mental self and our physical self? This lecture considers a range of phenomena and a handful of theories suggesting that those two problems are ultimately artificial. They are problems of our own making.

Neglected Areas

- In the previous lecture, we took a tour of the brain. The brain functions only as part of the larger nervous system. In our search for consciousness we neglected a number of structures crucial to the autonomic nervous system, generally thought of as unconscious.

- These include reflex structures within the spinal cord itself and the medulla, pons, and hypothalamus in the brain stem. These structures register and regulate blood sugar, oxygen, and carbon dioxide. They control blood pressure, heart rate, and body temperature: all the elements of homeostasis.

- It is the wider nervous system that leads people to seek food, water, sex, shelter, and warmth. It is also that wider nervous system that regulates sleep, dreaming, wakefulness, and attention.

- There are actually two subsystems of the autonomic nervous system: the sympathetic nervous system, functioning from the middle stretch of your spinal cord, and the parasympathetic, functioning from the upper spinal cord and brain stem.

- The sympathetic nervous system has been roughly characterized as an initiating system and the parasympathetic as an inhibitory system. The sympathetic accelerates the heartbeat. The parasympathetic slows it.

- There is also a third nervous system embedded in the gut, sometimes described as a "second brain." It's the enteric system, a mesh of neurons that regulate your gastrointestinal system. Although it normally communicates with the sympathetic and parasympathetic systems, the enteric system is autonomous: It will continue to function on its own even when those connections are cut.

The Mind

- The brain isn't separated from the larger nervous system. So why think the mind is separated from the body? After all, what would it be like if your mind was separated from your body?

- You couldn't feel pain, for one thing. That might sound wonderful, but it's not. A man named Steven Pete has a congenital inability to feel pain. His parents realized the problem when their toddler chewed half of his tongue to ribbons. Those with the disability are prone to bone fractures they don't realize, problems from eye irritations they don't feel, cavity-caused toothaches that don't ache, and infections for which they can't feel the symptoms.

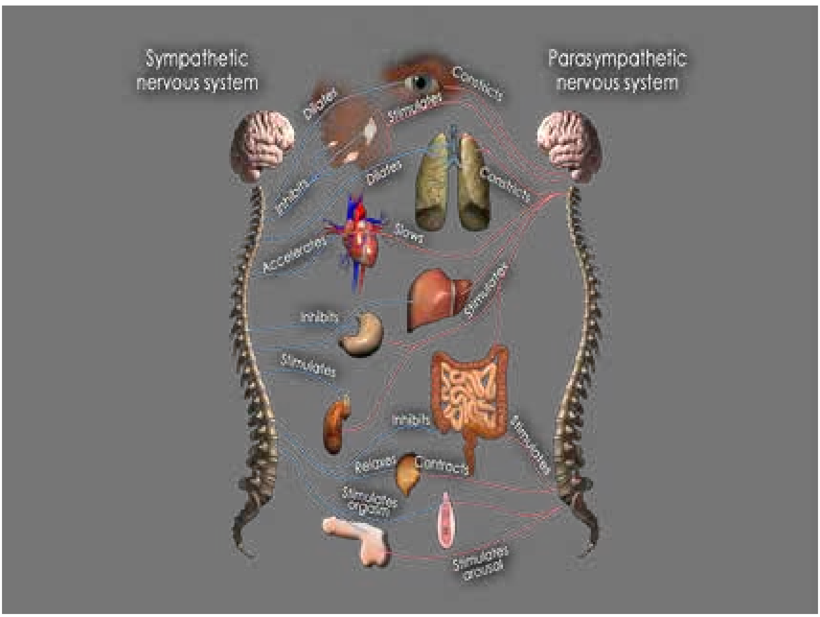
- Congenital inability to feel pain is a condition researchers were surprised to find was linked to mutations in a single gene: SCN9A, which provides instructions for an element crucial to the proper functioning of the nerves that transmit pain signals.

- If your mind were separated from your body, you also wouldn't be able to feel the position or movement of your body. People have an internal sense of their own body called proprioception; you can test yours by closing your eyes and touching your knee. It should be no problem.

- But there are six people in the world known to have lost all proprioception. Ian Waterman is the only one who has successfully compensated for it. He makes up for the loss of proprioception by using vision. He has to monitor his body visually, constantly, in order to move with anything like a semblance of normality.

Thinking

- We tend to think of thinking—of cognitive processing—as something that happens in the head. But maybe we use our bodies for at least some of our cognitive processing. For instance, is a child counting on her fingers doing all her thinking inside her head?



- There are films of Walt Disney's animators drawing sketches of characters with strong emotions: anger, fear, surprise. The animators contort their faces into those expressions in order to draw them.

- When thinking internally—with "inner speech"—people's mouth and throat muscles make minute movements. This is called subvocalization and shows that even silent soliloquies aren't totally private.

Psychological Theory

- There is a tradition of psychological theory, widely applied in a number of fields, that emphasizes both the mind in the body and the mind in the world.

- J. J. Gibson was an American psychologist. Working from the 1950s through the 1970s, he offered a new theory of perception. The standard view at the time was information-processing theory in terms of representations. The idea was that, in perception, the mind builds representations of objects in the world around it.

- For Gibson, that was a picture of a mind separated from its world. The core of Gibson's theory of perception is that we don't perceive objects and don't operate cognitively in terms of object representations. What we perceive—what any animal perceives—are what Gibson terms affordances.

- Squirrels don't see trees, represent them internally, and calculate how to climb them. What they see is something more immediate and more action-oriented than that. They see a way up. For Gibson, a mind in the world operates in terms of those affordances: more immediate and more action-oriented than mere representation of objects.

- The theory has had a wide impact. A range of thinkers in both philosophy and psychology carry on the Gibson tradition, emphasizing an embodied mind interactive with its environment. Called situated cognition, the theory is often cited as background for "knowing by doing." It forms the basis of participatory learning strategies in educational psychology.

- Critiques of this approach warn against is overstating the conclusions. It may be true that we learn by doing without it being true that it is the only way we learn.

Evolutionary Pressures

- If we look at the issue from an evolutionary perspective, it shouldn't be surprising that we find important areas of integration between mind and body, and between mind and world.

- Evolutionary pressures operate on organisms in environments, not on minds separate from bodies or minds independent of the world. Successful organisms are those that reproductively succeed. A successful organism will have mental responsiveness reflected in environmentally sensitive bodily behavior.

- We are the product of that kind of evolution. Despite a philosophical history of conceptual separation, we shouldn't be surprised that when we look closely at our own experience, we find a mind integrated with its body and situated in its world.

Robotics

- That evolutionary perspective on mind and body is reflected in robotics. Rodney Brooks is a robotics professor at MIT. He is known for solving problems by taking unusual approaches.

- Brooks proposes building robotic brains in layers, just like evolution seemed to build our brain: Start with a basic navigational unit, then add different perceptual units, then add something like a goal-directing unit, and so on.

- His robots don't add a body to an established artificial intelligence program. They operate with a robotic body and an interactive environment from the start. His robots are made to scramble over terrain. They end up looking and moving like insects because that is what proves effective in that environment.

Mental Concepts

- Everyday experience, situated cognition and affordance theories, and even robotics argue for paying attention to connections between mind and body. Even our concepts of purely mental things have lots of the world built in.

- The philosopher Hilary Putnam argues that even meanings aren't purely contents of the head. The meanings of your words don't depend merely on how your linguistic community uses them. The meanings of your words can depend on the world itself.

- This is known as Putnam's Twin Earth thought experiment: What does the word *water* mean? Water is H_20, two hydrogen atoms bonded with one oxygen atom. Now, suppose we send a group of astronauts to another planet, one very much like ours. The planet is Twin Earth and in most respects is very similar to our world.

- But the stuff that fills the streams and rivers on Twin Earth, the stuff they pour into glasses and drink, has a very different chemical composition. It has a complicated formula abbreviated as XYZ.

- Suppose you have a doppelganger on Twin Earth, and suppose that you sit down at a restaurant and ask for a glass of water. To you, water means H_2O. But on Twin Earth, your doppelganger asking for a glass of water would be asking for XYZ.

- Your psychological states—what is in your heads—are precisely the same. But the meaning of the word *water* as you use it and the meaning of the word *water* as your doppelganger uses it are different. Here, for you, it means H_2O. There, for your doppelganger, it means XYZ. Putnam concludes that the meanings differ, even though everything in the heads is the same.

- Putnam's position is called externalism: When we talk about psychological states like meaning, we aren't just talking about something internal. Whether someone can be said to mean something may depend on how the world is.

- The question is how far to push the theory. A common response to externalism is to concede that some psychological terms, in some uses, may indeed fit Putnam's story. Some uses of belief and meaning may have wide content, meaning that they depend on the world around us. But that may not be true of all our psychological or mental terms. Other psychological concepts may have narrow content, limited more narrowly to just what's in a person's head.

Further Externalism

- The philosophers Andy Clark and David Chalmers argue for an extended-mind thesis that they call active externalism. In cases in which we use the world as part of our cognition, we should think of our minds as extending into the world. Whatever you use to think is part of your mind.

- They use this comparison: Igna and Otto both want to go to a new exhibit at the Museum of Modern Art. Igna recalls the museum's location—it's on 53rd Street—from memory and starts walking there. Otto, who has Alzheimer's, consults the notebook he carries and brings up the museum's location that way. He also heads to the museum.

- Clark and Chalmers say that both Inga and Otto believe that the museum is on 53rd street because both have that as accessible information. Where it is accessible, whether in skull or notebook, shouldn't make a difference.

- Clark and Chalmers concede that we don't use mental terms that way. We say Inga remembers where the museum is but Otto doesn't. He has Alzheimer's. It's because he can't remember things that he has to consult his notebook. Inga has a background belief that the museum is on 53rd street. Otto just has a jotting in his notebook.

- But even if we don't use mental terms that way, Clark and Chalmers claim we should. In all important respects, the cases of Otto and Inga are relevantly the same. If our mental concepts don't yet recognize a mind extended into the world around it, they should.

- Taken that way, they aren't just giving us an analysis of our current concepts of mind. They're telling us to change them. Is that going too far?

Suggested Reading

Clark and Chalmers, "The Extended Mind."

Gibson, "The Theory of Affordances."

Questions to Consider

1. Give three examples from your own experience in which you use your body as part of your thought process.

2. Give three examples in which you use something from the world as part of your thought process.

Lecture 19

Francis Crick and Binding in the Brain

The central figure in this lecture is Francis Crick, co-discoverer with James D. Watson of the structure of DNA. But his work on DNA is not the main topic this lecture will discuss. Instead, the lecture concentrates on his later work on the brain, specifically on two of Crick's hypotheses regarding brain function and consciousness. The first hypothesis is that the theory is that patterns of neurons firing synchronously in the range of 40 hertz form part of the same conscious experience. In the second, Crick proposes that there is a particular part of the brain that is responsible for the coordination across various areas of the brain required for consciousness.

Francis Crick

- After their work together on DNA, Crick and Watson went on to important later careers. Watson's was the straighter line: He went to the biology faculty at Harvard and from there to direct the Cold Spring Harbor Laboratory on Long Island. By the 1990s, he was a major player in the Human Genome Project.

- By contrast, Crick left DNA behind and changed course again. He felt there were two mysteries: the physical basis of life and the physical basis of mind. His progress on the question of life had been astounding. So Crick turned to the second question: What is the physical basis of mind? What is the underlying structure of consciousness?

- Crick was a materialist, or physicalist, with regard to mind as he was with regard to life. He was sure the underlying structure of consciousness was going to be a physical understructure.

- In 1990, Crick published a groundbreaking piece with Christof Koch, a neuroscientist. It's called "Toward a Neurobiological Theory of Consciousness." At the beginning of that piece, they were able to say, "It is remarkable that most of the work in both cognitive science and the neurosciences makes no reference to consciousness." This was a piece that changed all that.

- Crick and Koch laid out the problem of consciousness as a genuinely scientific project. They lay out a number of assumptions they're going to make, as well as a number of questions they're going to put aside, at least at the beginning.

- Their first assumption is that it's not just people who have consciousness. Other animals, and clearly the higher mammals, do as well. They assume therefore that language of the sort found in humans is not necessary for consciousness.

- One of the problems they put aside is how far down an evolutionary ladder consciousness goes. They claim it's unprofitable at an early stage to speculate as to whether octopi, fruit flies, or nematodes are conscious. They also put aside, for the moment, questions of what consciousness is for.

- There are two aspects that they put aside that may be particularly worrisome. One is what they call the problem of qualia, focusing on the character of subjective experience. The other issue they put to the side is actually defining the phenomenon they are after.

- Those are some pretty significant factors to set aside, but it's all in the name of trying to convert the problem of consciousness into a genuinely scientific research project. They recognize their first step as precisely that: a first step. In that regard—jumpstarting a scientific study of consciousness—the Crick and Koch paper was a dramatic success.

- That central aspect of consciousness is what they term the binding problem. Although they don't allude to it, the binding problem has a significant philosophical history.

The Binding Problem

- The binding problem is this: At the moment there are certain things you see. There are certain things you hear. You may feel certain pressure on your fingertips or a certain tension in your forehead. There are certain thoughts going through your mind, and you might smell coffee.

- Those are all different sensations and conceptualizations. But they are all happening at once, and they are all happening together. Yours is a single consciousness in which many things are happening simultaneously. How precisely is that possible, that all those things are bound in a single consciousness?

- The binding problem can be phrased as either a problem of phenomenal binding—what brings all those phenomena together?—or of binding in the brain.

- Crick and Koch approach it as a problem of the brain. We know that different parts of the brain process vision, hearing, touch, smell, and thoughts. What binds activity in those different parts of the brain together into a single consciousness?

- Crick and Koch weren't the first theorists to ponder the binding problem. It is raised in both Plato and Aristotle. And the Scottish philosopher David Hume, writing in 1738, raised the binding problem in a particularly forceful way. Hume was an empiricist. The basic idea of empiricism is that all knowledge comes from experience.

- The mind operates using ideas from impressions as its basic material, but there are important limitations as to what it can do with them. In Hume, association of ideas operates solely in terms of similarity, whether things are next to each other in time or space, and an association of cause and effect that arises purely from habit.

- How are these impressions bound together? In the main section of Hume's *Treatise of Human Nature* in which he addresses the issue, he simply denies that experience gives us any sense of self. If ideas come from impressions, there is no genuine idea of self. What are people? All of mankind,

Hume says, "are nothing but a bundle or collection of different perceptions, which succeed each other with an inconceivable rapidity, and are in a perpetual flux and movement."

- It's called the bundle theory precisely because of that passage. The bundle theory doesn't so much solve the binding problem as refuse to recognize it. Nothing brings experiences and sensations together. They remain separate. They're as loose as a bundle of twigs. But in a later appendix to the *Treatise of Human Nature*, however, Hume confesses that he's not satisfied with his earlier analysis.

- Writing 40 years later, near the end of the 18th century, Immanuel Kant credits Hume with awakening him from his dogmatic slumbers. One of the problems in Hume that Kant tackled was the binding problem.

- For Kant, the mind has templates in advance that shape all the experience that comes in. One of the templates for Kant is space. Another is time. Space and time aren't so much in the world as in how we come prepared to interpret the world.

- The same goes for a concept of self and a binding of experience into a single unity. Kant has a fancy name for it: the transcendental unity of a perception. "Unity" signals it's binding; "apperception" signals it's a kind of perception or representation; and "transcendental" signals it doesn't come from experience but is necessary for experience.

Back to Crick and Koch

- Crick and Koch see attention and short-term memory as crucial for consciousness, but those don't solve the binding problem. As Crick and Koch admit, there is no evidence of any spatial location in the brain in which "it all comes together." What they propose instead is a temporal binding: a binding in time rather than space.

- How do different areas of the brain bind in consciousness? Their hypothesis is that binding occurs through the synchronization in the oscillation patterns of

neuron firing commonly called brain waves. EEG recordings show particular firing patterns in particular areas of the brain. Sometimes these are slow—alpha waves at 8–12 cycles per second, or 8–12 hertz. Sometimes these are fast—gamma waves, at 40–70 hertz.

- Crick and Koch cite experimental research on cat brains showing that when a cat is shown a moving bar, neurons some distance apart in its brain respond with synchronized oscillations. They propose that is what all binding is, and ultimately what consciousness is: "We suggest that one of the functions of consciousness is to present the result of various underlying computations and that this involves an attentional mechanism that temporarily binds the relevant neurons together by synchronizing their spikes in 40Hz oscillation."

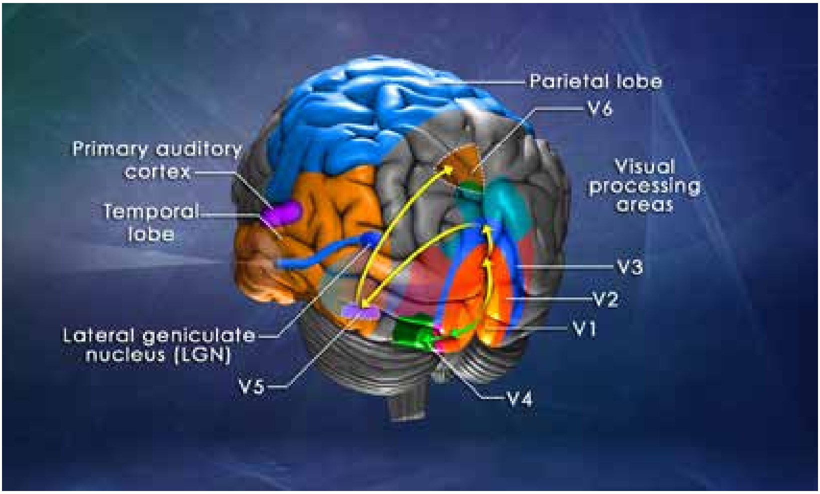
- That is Crick's first hypothesis. It had an immense influence precisely because it was a serious scientific hypothesis about consciousness. It opened the door to a field of research that has flourished ever since.

- Summed up, the theory is that patterns of neurons firing synchronously in the range of 40 hertz form part of the same conscious experience. Those firing asynchronously, or out of that range, may well be functioning unconsciously, but don't break the surface of conscious awareness.

- How has Crick's first hypothesis fared? A number of studies, in both animals and humans, do indicate some forms of synchrony as important for binding different perceptions. However, it does not appear that the importance of synchrony is limited to the 40-hertz range, nor that it needs to have the form of regular oscillation that Crick and Koch emphasize. On the other hand, there are cases of 40-hertz synchrony that do not trigger conscious awareness.

The Second Hypothesis

- By 2003, research had led Crick and Koch to abandon the first hypothesis, at least in its original form. They wrote, "We no longer think that synchronized firing, such as the so-called 40 Hz oscillations, is a sufficient condition for" consciousness. Nor did they appear to think that it is necessary for consciousness.



- But Crick didn't give up. He was working on a paper that lays out a second hypothesis regarding consciousness up to the day he died in 2004. In his second hypothesis, Crick targets a particular anatomical area. Crick proposes that there is a particular part of the brain that is responsible for the coordination across various areas of the brain required for consciousness.

- The title of the piece Crick was working on when he died is "What Is the Function of the Claustrum?" The hypothesis is that it is the claustrum that is crucial to consciousness by binding the different areas of the brain.

- *Claustrum* literally means "hidden away," and the term is apt. On each side of the brain is a lateral sulcus, a deep canyon that separates the temporal lobe from the parietal and frontal lobes. Deep in that canyon is a folded part of the cortex called the insula. One step even deeper toward the center of your brain is an extremely thin and irregular sheet of neurons, hidden away. That's the claustrum.

- Crick's work suggests that the claustrum coordinates different areas of the brain, binding them into consciousness. The brain may not have a spatial theater where everything comes together. But Crick suggests that it does have an anatomical feature, spatially located, that coordinates the temporal binding of firing in different parts of the brain.

- Crick proposed a test of his hypothesis: Knock out the claustrum and see what happens. He also noted the extreme difficulty of performing any such test. The claustrum is so thin that shutting it down, and only it, either pharmaceutically or surgically seemed beyond the reach of current techniques.

- Much more recently, however, Mohamad Koubeissi of George Washington University and his collaborators in Marseille and Geneva have reported a suggestive study. They planted individual electrodes in different parts of the brain of a 54-year-old woman with intractable epilepsy. Koubeissi's attempt was to treat the woman's seizures by electrical stimulation at specific points in the brain, tracking the effects on synchronization at they did so.

- Stimulation of a single electrode implanted in many parts of the brain seemed to have no discernible effect on synchronization. An electrode in the hippocampus, amygdala, and many parts of the neocortex could be activated without significant effect.

- Results were dramatically different for an electrode stimulating one particular point in the brain. Described as a "consciousness switch," electrical stimulation at that point seemed to disrupt firing across the brain in a way that rendered the patient entirely unconscious. Where was it? Right next to the claustrum.

- There is also some evidence against the claustrum hypothesis. The structure of the claustrum seems too simple for any complex information processing. Brain imaging doesn't indicate that it is particularly active during consciousness.

- Whether it ends up being right or not, Crick's second hypothesis is another characteristically bold speculation as to how the brain works. Like his first hypothesis, the second is a clear conjecture framed in light of available evidence and fully open for further testing. In that alone, it continues to guide research and advance the field.

Suggested Reading

Crick, *The Astonishing Hypothesis*.

Crick and Koch, "What Is the Function of the Claustrum?"

Watson, *The Double Helix*.

Questions to Consider

1. Has the course of your life been a straight line between points A and B, or something with more turns and twists than that? Which parts have been the most interesting?

2. This lecture focuses on the binding problem, in two forms:

    - The problem of brain function: What binds the functioning of different parts of the brain?

    - The phenomenal problem: How do all your different sensations come together in a single consciousness at a single time?

3. In what respects do you think those are the same question? In what ways are they different questions?

Lecture 20

Clues on Consciousness from Anesthesiology

In the city of Boston stands a 40-foot-tall monument dedicated to the use of ether as an anesthetic. At the top is a vaguely Moorish figure in robe and turban holding the drooping body of a man over his knee and a cloth in one hand—a cloth presumably soaked with ether. The innovation of anesthetic ether was a major achievement in medicine. The American dentist William T. G. Morton and the Harvard professor Dr. Charles Jackson squabbled over credit for the invention. Whoever deserves the credit, the use of ether spread quickly, as did other anesthetics, like chloroform and nitrous oxide. Their use has much to teach us about consciousness.

On Anesthetics

- The initial hypothesis on different anesthetics was that they must have some common mechanism. There must be some one thing that all these different chemicals are doing in the brain.

- In 1901, a couple of independent studies reported that there was a correlation between the effectiveness of a chemical as an anesthetic and whether it was soluble in olive oil. Lipids are fats, so what emerged was the lipid hypothesis of general anesthesia. All of these chemicals must work by affecting lipid membranes in the nerves.

- By the 1980s, the lipid hypothesis had been debunked. An even larger range of anesthetics was available by that time. But how those different anesthetics worked was still a mystery.

- By then, attention had turned away from lipids to proteins and ion channels instead. Theories also shifted away from a unitary hypothesis to a diversity hypothesis: Maybe all anesthetics don't work the same way.

- It is also possible that the effects aren't even all that similar. *Anesthesia* is a single term meaning "insensitivity to pain." But there are more terms beyond that.
    - *Paralytics* is the term used for chemicals employed as muscle relaxants, which block patient movement.

    - *Amnesiacs* is the term for chemicals that block recall. What happened to you for a period of time after you took an amnesiac won't be remembered afterward.

    - *Analgesics*, like aspirin and acetaminophen, dampen or deaden pain without loss of awareness.

- The aim of general anesthesia is to wipe the awareness slate clean—to entirely eliminate conscious awareness, eliminating pain in the process. A warning: If you have a surgical procedure scheduled in the near future, you may want to postpone this lecture until later.

The Effects

- It is common in anesthesiology to use a cocktail of chemicals that have a spectrum of effects. The combination typically includes a paralytic for muscle relaxation and smooth surgery. It also typically works as an amnesiac to block memory of the event. Of course, the main intent is to block pain and awareness under the knife.

- A philosophical problem arises immediately. Suppose, for example, that you have a combination that is very effective as a paralytic (the patient doesn't thrash about) and is also very effective as an amnesiac (the patient doesn't complain after). Isn't it still possible that the patient felt pain—perhaps a great deal of pain—during the operation?

- Indeed, the awful truth is that this apparently happens. In 1984, four national newspapers in Great Britain carried the following short advertisement asking if people had been conscious during surgery. They got all too many responses recounting terrifying experiences from helpless patients.

- The horror is called intraoperative awareness or anesthesia awareness. Reports of anesthesia awareness occur in only one or two cases in a thousand, but you sure don't want to be one of those unlucky ones.

- Moreover, the figure of one or two cases in a thousand represents only those cases in which patients remember being awake and aware; an effective amnesiac may simply erase the memory of the horror.

- What we need is a consciousness monitor. With that, we could tell that the patient is really unconscious when under anesthetic, not merely paralyzed with muscle relaxant and unable to recall the physical experience later.

The Four Questions

- Four questions need to be answered before a consciousness monitor is possible. By now the first question should be familiar: What is consciousness? The primary answer seems to be subjective experience. We want to eliminate the subjective awareness reported in the horror stories. The experience that we want the patient to have during the surgery is none at all.

- The second question should also be familiar: How does consciousness work? If we could answer that, we would know the physical process we were trying to turn off in trying to turn off consciousness.

- A third question concerns anesthetics in particular: How do anesthetics work?

- A fourth question is a question of measurement: Can we measure that anesthetics are in fact doing that work? It's at that stage, with answers to all four questions, that we will really have a consciousness monitor.

Research

- The last few years have seen a number of promising breakthroughs by a number of different research teams. George Mashour, an anesthesiologist at the University of Michigan, is one of the main players. He speaks of an explosion of studies on how anesthetics interrupt consciousness and what they can teach us about it.

- Before the 1980s, the idea was that different anesthetics must work in basically the same way—the unitary hypothesis. In the 1980s, thinking shifted to a realization that different anesthetics or different aspects of anesthetics may work in different ways.

- In the past few decades, research has swung back toward a unitary hypothesis of consciousness once again. With that shift has come new hope for answering some of the questions we've highlighted.

- What made a unitary theory seem impossible was that different anesthetic agents seemed to do very different things to very different parts of the brain. Rather than a general theory of how anesthesia works, it appeared that the best we could get was a bunch of theories of how a bunch of different anesthetic agents work.

- This overlooks a simple distinction between what is done and the different ways in which that thing is done. It is possible that the same thing can be accomplished in a number of different ways.

- Maybe anesthetics have the same general function, producing the same overall effect, although the details of how they accomplish that general function are different. Our unitary theory of anesthesia will be a theory of the general mechanism that produces the general effect.

- Many theories involve approaches that take binding to be crucial to consciousness. The idea is that subjective experience occurs not in some single part of the brain but in the operation of multiple parts of the brain functioning together.

- If we accept a binding hypothesis, we have provisional answers to the first two of the four questions. And with those, we can propose a possible answer to the third.
    1. What is consciousness? The sense of consciousness that seems to be important for studies of anesthesia is subjective awareness.

    2. How does consciousness work? Consciousness operates any of various forms of the binding hypotheses. It is binding across a brain that is necessary for consciousness.

    3. How do anesthetics work? Perhaps by blocking or interfering with the binding necessary for consciousness.

Studies

- A number of studies across the world seem to support the combination of a binding hypothesis regarding consciousness and this unitary hypothesis of how anesthetics work.

- A Belgian group has studied anesthesia with PET and functional MRI scans. They found that when patients are unconscious, external stimulation causes initial islands of activation to occur, but it doesn't spread to other areas of the brain like normal.

- A German team has studied the process in slow motion by altering how anesthetics are administered; they administer the drug Propofol so that patients fall asleep in minutes instead of the normal 10 seconds. During the administering, they observe what happens in EEG readings with a mild shock to the volunteer's wrist. When awake, the result is activation of the sensory cortex, with further activation of frontal lobes associated with judgment and temporal lobes associated with memory. When fully sedated, the activation seems to stop at the sensory cortex.

- This study resonates with the unitary hypothesis. They support the idea that what anesthetics do is block the process of binding across the brain.

Blocking

- How precisely is binding blocked? There are some intriguing research findings here as well. When conscious, the brain shows active communication between importantly different patterns of activity in different parts of the brain.

- As the patient slips into unconsciousness, that active exchange between areas of differential activity is replaced with a strong, monotonously synchronized, slow oscillation that washes across them all.

- A number of researchers have proposed that this is the process of "unbinding" in the brain. Andreas Engel's results indicate that communication between areas in the brain is shut down not with a sudden silence between them but rather by the imposition of that abnormally strong synchronization.

- Like all good science, these results are open to further testing, vulnerable to falsification, subject to replication, and open to refinement. There are more, both older and newer, that seem to track the same trajectory. They offer strong evidence that anesthesia works by interfering with communication across different areas of the brain.

- We know a great deal more about how anesthesia works than we did only a few decades ago. That in turn gives us clues to at least some aspects of the neural dynamics responsible for consciousness.

- Regarding the fourth question: Do these results give us what we need for a consciousness monitor? If binding is a necessary condition for consciousness in the sense we're after, and if we know anesthesia operates by blocking that binding in measurable ways, it looks like we might actually be able to build such a device.

Philosophical Questions

- On the assumption of a binding theory of consciousness, we've sketched some promising answers to the four practical and scientific questions. But major philosophical questions remain.

- Even the most successful binding theory of consciousness would leave a crucial question unanswered: Why is it that subjective experience emerges only with binding across a brain?

- There are also some major philosophical assumptions we have made in tracking the promise of scientific breakthroughs in anesthesiology. We can still distinguish between two things, at least conceptually. On the one hand: a genuine anesthetic, genuinely eliminating awareness and pain during an operation. On the other: a highly effective combination of a paralytic together with an amnesiac.

- In the second case, pain and awareness are there during the operation, though the patient is disabled from crying out and the memory of the event is entirely wiped afterward.

- Neither an outsider's observation nor any response from the patient at any time would tell us if the patient felt sensation during the operation: We hear no complaint when he is paralyzed during the operation, and hear no complaint afterward because he has forgotten all about it.

- That philosophical possibility would remain even if we did have the kind of consciousness monitor that we've outlined. We would know that when the

fMRI indicates that functioning is confined to specific areas of the brain, or EEG readings signal the synchronous wave across the brain, the horror stories stop. We would no longer hear patient reports of anesthetic awareness.

- But it would remain a philosophical possibility that we no longer hear reports merely because we've finally found brain states from which we can't hear the screams: brain states from which all reporting is blocked.

- Perhaps it's *merely* a philosophical possibility. Perhaps it will join the wild speculative possibility that Bertrand Russell envisages: the possibility that the universe was created five seconds ago, complete with false memories. Or the logical possibility that other people are merely clever robots. Those too are philosophical nightmares. Luckily, blissfully, they prove unsustainable in the normal course of everyday life. Maybe that's where this possibility belongs as well.

Suggested Reading

Mashour, "Integrating the Science of Consciousness and Anesthesia."

Mashour and Alkire, "Evolution of Consciousness."

Questions to Consider

1. A worry is raised in this lecture that there may be cases in which a patient is fully conscious but paralyzed during an operation, but the horrors are not reported later because the cocktail of drugs administered also includes an amnesiac that wipes memory of the event. If the patient doesn't remember the event later, is it nonetheless something we should worry about? Why?

2. It has been suggested that anesthesia, vegetative coma, and certain stages of sleep all show a pattern of unbinding in the brain. From your own experience of sleep, what do you think the pattern of unbinding is? Is it gradual, or sudden?


12

Lecture 21

Of Mind, Materialism, and Zombies

This lecture discusses zombies—not the blood-dripping creatures of popular portrayals on large and small screens, but denizens of philosophical thought experiments. These zombies don't eat brains. They're meant to stimulate brains: to test philosophical intuitions and theories about minds, brains, and consciousness. A zombie is a creature just like a person: acting like a person and saying the things a person does. The only difference is that a zombie does it all without any consciousness. A zombie's responses are identical but aren't routed through subjective experience the way a person's are.

Materialism

- Materialism is the view that everything is ultimately physical. Typically, we picture the physical universe as a hierarchy: one type of physical thing is composed of another type of physical thing.

- For instance, you are composed of complex organs, which are composed of tissues, which are composed of cells, which are built from proteins, and so on, down to atoms and subatomic particles. Hierarchies can also be traced from the bottom up.

- That hierarchy of things corresponds to a hierarchy of sciences. At the top, we have sociology and anthropology—sciences of societies and cultures—

and economics—a science of a particular dynamics within societies. Societies are composed of people, the realm of social and individual psychology.

- People, in turn, are biological organisms, with various branches of biology on the next level of the scientific hierarchy. Biology operates in terms of organic chemistry, the next rung down on the ladder. Below chemistry is physics, leading us to a foundation in particle and subatomic physics.

- Reductionism is a theory in philosophy of science tied to that ladder image of scientific disciplines. The theory is that each step in the ladder reduces to the step below. The reductionist view is that the foundational science is physics. If and when we get the whole picture, we'll see that all other sciences are merely branches or sub-disciplines of physics. All science ultimately reduces to physics.

A Hard Look at Reductionism

- Although the hierarchical picture is appealing, what do we actually mean when we say that one theory reduces to another? Here is the classical answer: The "higher" theory (chemistry, for instance) reduces to the "lower" theory (physics) if all the laws of the higher theory can be deduced from the lower theory.

- Some might object, citing that chemistry and physics don't even use the same terms. Classical reductionism has an answer to this: Reduction consists of deduction plus definition. We simply need to add bridge principles which define the concepts of the higher theory in terms of the concepts of the lower.

- That grand reductionist scheme fired the imagination of philosophers of science well into the middle of the 20th century. But over time, the difficulties facing the reductionist vision became clear. For instance, if we want psychology to give us an understanding a thing of mental illness, classical reductionism would make that difficult: How precisely would one define the concept of psychosis in terms of subatomic particles? And how much understanding of mental illness would that actually provide?

- The crucial point is that the abstract models we build to understand different levels of phenomena are inevitably going to be different. The concepts we take as basic, the relationships we choose to model, and the dynamics we put at the center will be different.

- In this course, we are most interested in the question of reduction between two particular areas: between the mental and the physical, from mind to brain, or perhaps from psychology to neurophysiology.

- There is a strong argument that reduction at this point will prove impossible. It's called the multiple instantiation argument. It goes like this: Suppose we found very different creatures on a distant planet—creatures based on silicon rather than carbon, with brains that were distributed across their bodies in ways we didn't recognize.

- These creatures could conceivably feel pain But that means that the concept of pain is multiply instantiable. There might be various kinds of creatures, with various kinds of nervous systems—even nervous systems totally unlike any we could imagine—who still feel pain.

- If so, we're never going to be able to define pain as a particular type of carbon-based brain function. It's a concept of a very different kind—a concept at a very different level of abstraction. If our mental concepts are multiply instantiable, that kind of reduction of mind to brain will inevitably fail.

Elements of Value

- The full classical picture of reduction is now generally recognized as a failure. However, if we let go of the full classical picture, there may be elements of reductionism that are still of value.

- The classical picture called for what is sometimes called type-to-type reduction. The idea was that types of mental states, like pain, would correspond to types of brain state. The multiple instantiation idea dashes that hope.

- But it might still be true that each particular instance of pain, in a particular creature at a particular time, corresponds to a particular brain state in that creature at that time. We can call that particular instance of pain in that creature at that time a specific token instance rather than a general type. Perhaps token-to-token reduction is possible even if we've abandoned type-to-type.

- Another valuable thing that we might take from the reductionist program is a simple research strategy. If you want to understand something, it's often a good idea to take it apart to see how it works.

Non-Reductive Materialism

- Reductionism is an epistemological theory: a theory about knowledge. Materialism, on the other hand, is an ontological theory. It's a theory of being: Everything that exists is ultimately physical. Reductive materialism insisted on putting those two things together. Another option is to pull them apart.

- People have tried to flesh out the possibility of a non-reductive materialism by borrowing an idea from the field of ethics. Suppose that there are two actions that happen in identical circumstances, by identical agents, with identical backgrounds, motivations, and consequences. Imagine something like identical murders happening in Chicago and Detroit.

- If all the circumstances are the same, could one of those actions be ethically right and the other be ethically wrong? No.

- In philosophy of mind, the corresponding proposal is that mental states depend on brain states in much the same way. Just as you couldn't have differences in right and wrong without some difference in circumstances, you couldn't have differences in mental states without some difference in underlying brain states.

- We'll call it dependence materialism. Mental states depend on brain states. Everything is still ultimately physical.

Zombies

- Zombies offer a thought experiment to test, on reflection, whether dependence materialism is plausible.

- Suppose that somewhere, in some faraway galaxy, you have a zombie twin. There is only one difference between the two of you: You have a rich inner life and are conscious.

- Being physically identical, your zombie twin does everything you do. But it's all merely a physical operation. Zombies have no subjective awareness and lack consciousness.

- Is that kind of zombie possible? People who say yes are anti-materialist. Even dependence materialism won't do. Materialism says that if the two of you are physically identical, the two of you must be mentally identical. People who say yes fall more in the camp of dualism.

- People who say no, such a zombie is impossible, fall firmly in the materialist camp. A physically identical but mentally different zombie would be impossible if the mental must depend on the physical.

More Zombies

- To begin with, there are different ways in which we might think of you and your zombie twin being "identical." So far we thought of you and your zombie twin being fully physically identical: identical atom for atom, molecule for molecule.

- Alternatively, we might think of a weaker kind of identity. Suppose we specify that you and your zombie twin are not physically identical, just behaviorally identical. Everything you do or say, the zombie does or says.

- Behaviorism appears in a number of different forms. One view is that mental states are dependent on behavioral states. If two things are behaviorally identical, they have to be mentally identical too.

- We can use a different kind of thought experiment zombie to test whether you're this kind of behaviorist: Could you and your zombie twin be behaviorally identical, but you be conscious and your zombie twin not be?

- If you think that kind of zombie is impossible, you're not merely a materialist, you're a behaviorist as well. If you think that kind of zombie is possible, you're an anti-behaviorist.

Types of Possibility

- Possibility comes in different strengths. Something is logically possibly if imagining it doesn't involve some kind of contradiction. For instance, imagining a horse with three horns would be weird, but doesn't involve a contradiction. A married bachelor, though, is a contradiction.

- On the other end of the possibility spectrum is a laws-of-nature possibility. To ask whether something is possible in this sense is to ask whether it's possible without violating the laws of nature in the actual universe. For instance, traveling faster than the speed of light is not something that could happen without violating the laws of nature that actually hold in our universe.

- We can add one more type of zombie. The multiple instantiation argument argued that any of various creatures, with all kinds of different anatomies, silicon-based or carbon-based, might all feel pain. Pain isn't something that can be reduced to a particular physical state, because it seems that very different creatures on other planets might all feel pain.

- Hilary Putnam is the philosopher who proposed that argument. He believed it shows that pain isn't a physical state; it's a functional state. Putnam's functionalism is a descendant of behaviorism, emphasizing a functional organization that includes internal states as well as behavior.

- Consider your functionalist zombie twin. There is no guarantee that the zombie is physically or behaviorally identical to you. The zombie has internal states that correlate with each of your mental states. Is it logically possible for you to have a functionally identical zombie twin like that, but for you to be conscious and your twin not to be?

- Are functional zombies laws-of-nature possible? They would be robots, just like the behavioral zombies were robots, but with a particular organization of states that match the functional organization of whatever person they are copying. Would the laws of nature dictate that they then had to be conscious?

- To track your answers to questions like these, fill out a zombie scorecard. See this lecture's Questions to Consider below for a model.

Suggested Reading

Dennett, "The Unimagined Preposterousness of Zombies."

Flanagan and Polger, "Zombies and the Function of Consciousness."

Polger, "Zombies Explained."

Questions to Consider

1 The multiple instantiation argument has us imagine very different creatures that nonetheless have the same mental states—pain, for example. Are there limits to that thought experiment for different mental states? What would a creature have to be like in order for it to have beliefs? To have a sense of pride?

2 Fill in your own zombie scorecard:

	Behaviorally Identical Zombie	Functionally Identical Zombie	Physically Identical Zombie
Logically Possible	☐ Possible ☐ Not possible	☐ Possible ☐ Not possible	☐ Possible ☐ Not possible
Laws-of-Nature Possible	☐ Possible ☐ Not possible	☐ Possible ☐ Not possible	☐ Possible ☐ Not possible

Lecture 22

Thought Experiments against Materialism

One of the main tools employed in contemporary philosophy of mind is the thought experiment. Virtually all of the contemporary arguments against materialism and functionalism rely on appeals to thought experiments. Many thought experiments, in philosophy as well as physics, are designed to show that a particular theory will generate a clearly unacceptable conclusion under clearly imaginable circumstances. In this lecture, we'll look at thought experiments in some detail.

Requirements of Thought Experiments

- The two requirements of thought experiments—a clearly unacceptable conclusion and clearly imaginable circumstances—are also potential weak points. A thought experiment from physics illustrates this point.

- Einstein wasn't a fan of quantum mechanics. He didn't like the essential and irreducible randomness at the core of the theory. It must be leaving out some hidden variables behind that apparent randomness.

- Working with colleagues Boris Podolsky and Nathan Rosen, Einstein came up with an argument against quantum mechanics. They designed a thought experiment to show that quantum mechanics couldn't be complete. Their strategy was to show that the target theory would lead to unacceptable conclusions under clearly imaginable circumstances.

- It's now known as the EPR paradox—EPR for Einstein, Podolsky, and Rosen. Here's the core: If the relevant interpretation of quantum mechanics were right, separating two particles that had once been entangled in interaction would allow information about a random event performed at one point in the universe to reach us at a different point in the universe instantaneously.

- If the target theory were right, it would then be possible for information to travel faster than the speed of light. But nothing can travel faster than the speed of light. If the relevant interpretation of quantum mechanics were right, we'd have "spooky action at a distance," Einstein said. This was an unacceptable conclusion.

- But the EPR thought experiment didn't turn out to be the knockout punch it was intended to be. There is ongoing work in physics attempting to convert that thought experiment into a real experiment. The underlying idea is maybe those unacceptable conclusions aren't really so unacceptable: Perhaps the universe doesn't have a speed limit when information is at stake. Perhaps information really could be transferred faster than the speed of light.

- Thought experiments often have that structure: In clearly imaginable circumstances, if the theory were right, you'd get an unacceptable conclusion, so the theory must be wrong. That can be a powerful argumentative structure. But it does rely on those two important points: The clearly imaginable circumstances really do have to be clearly imaginable, and the unacceptable conclusion really does have to be unacceptable. In the case of the EPR experiment, it's an open question whether that conclusion is really so unacceptable after all.

Frank Jackson

- The Australian philosopher Frank Jackson offers what has become one of the most famous thought experiments against materialism. The target theory is: If the universe is entirely physical, all the facts of the universe are physical facts.

- The clearly imaginable situation posits a brilliant neuroscientist named Mary. Mary grows up in a black-and-white room. She's educated through black-and-white books with black-and-white illustrations and views lectures in black and white.

- But since she's brilliant, she learns everything there is to know about the physical nature of the world, and in particular about the physical nature of color vision. She's trapped in a black-and-white world but becomes the world's expert on all the physical facts of color.

- One day, Mary is released from her black-and-white world, opens the door, and sees color for the first time: azure rivers, green fields, and so on. "Oh," she says, "So that's what color looks like."

- That's the clearly imaginable situation of the thought experiment. Here's the punchline: Hasn't Mary learned something new when she opens the door? We assumed that Mary knew all the physical facts of color before she opened that door. If she knows something more about color now, there must be something to know about color that isn't a part of mere physical knowledge.

- Materialism would entail that Mary already knew all the facts, and so couldn't have learned anything new. That is an unacceptable conclusion, so materialism must be wrong.

Leibniz and Dualism

- Another thought experiment comes from the early 1700s, with the German philosopher Gottfried Wilhelm Leibniz arguing for a very dualist conclusion. In the passage where this appears, Leibniz is arguing that perception and sensations are "inexplicable by mechanical causes." No machine, nothing that worked purely by mechanical means, could have perceptions or sensations.

- Suppose there were a machine whose structure produced thought, and sensation, and perception. Leibniz argues that's impossible. Imagine a mechanical head blown up big enough that we could walk inside. All the levers, gears, and pulleys are the same as before, only big enough that we can

walk among them. If the machine is really producing thoughts, sensations, or perceptions, it shouldn't matter how big we make the machine: All the causal interactions will be the same.

- Leibniz says inside the head, you'd see levers and gears and pulleys in motion. But you wouldn't see perceptions, or sensations, or thoughts. Leibniz argues that thought, sensation, perception, and subjective experience must therefore be something beyond the mechanism.

German philosopher Gottfried Wilhelm Leibniz

Criticizing the Experiments

- How good is the Leibniz thought experiment against mechanism or the Mary thought experiment against materialism?

- Leibniz's historical target was mechanism. Today it would be materialism, which argues subjective experience is to be explained in terms of a physical substrate. For those inclined to either position, Leibniz's thought experiment is far from convincing.

- Suppose that a machine in Leibniz's sense—he was thinking of metal clockwork—could produce sensations and perceptions. Why think that those would be additional things visible when we walked into it?

- Why not think that those things would be in the dynamics of the gears, pulleys, and wheels? They may be understandable once we consider the complexity of the system even if they're not something additional that we'd see on the tour. For instance, think of all the wonderful things a computer can do—spell checking, mathematical calculations—that we wouldn't be able to directly observe if we looked inside.

- Note that a materialist using the computer analogy to criticize Leibniz's argument hasn't thereby given an argument for materialism. But if a mechanist or materialist can appeal to that counter-Leibniz possibility, they can neutralize the threat of Leibniz's thought experiment.

- Regarding black-and-white Mary: Materialists have attacked Jackson's thought experiment at a couple of different points. The clearest counter-argument goes like this: We are supposed to buy the claim that Mary learns all the physical facts about color vision in a black and white room.

- But why think that? Why think that how red looks isn't a physical fact—a physical fact that, unfortunately, Mary couldn't learn in a black and white room? Why think that all facts can be expressed in black and white?

- That attacks the clearly imaginable part of the thought experiment. Maybe it isn't clearly imaginable that Mary learned all the facts about color vision in a black-and-white room. If that picture doesn't hold up, the thought experiment fails at step one.

Two More Thought Experiments

- Neither physical experiments nor thought experiments are unassailable arguments, invincible and beyond reproach. Let's consider two more thought experiments.

- The first is John Searle's Chinese room. Searle was a student of students of Wittgenstein. He offers a thought experiment against forms of artificial intelligence, part of Turing's legacy.

- Searle is arguing against strong artificial intelligence, a position close to both materialism and functionalism. The target theory is the claim that a computer running a program could understand a natural language. Could a computer running a program do what a human does when a human understands a language?

- Here is Searle's easily imaginable situation: He sits alone in a room with a slot in the door. Chinese speakers on the other side of the door write out questions in Chinese and slip them through the slot. Inside the room, Searle takes the input and follows a set of procedures outlined in a book written in English.

- The book was so good that the Chinese questioners become convinced there is a fluent (though slow) Chinese speaker on the other side of the door. The Chinese speakers may be convinced, but that's where Searle's argument starts.

- Searle's argument is essentially that he is taking input (the questions) and giving output (the answers) without understanding Chinese. The activity in the Chinese room parallels the operation of a computer running a program: input, rule following, output. If there's no reason to think that the Chinese room Searle understands Chinese, there's no reason to think that a computer running a program understands Chinese.

- Searle's experiment has had its own critics. Some of the critiques accept the thought experiment, but attack the conclusions that Searle thinks we should draw from it.

- Searle is not a dualist. He thinks that consciousness is a phenomenon in the physical brain. But the Chinese room seems to show that instantiating a program isn't enough. So what more do you need?

- The philosopher Ned Block has pressed the question with a thought experiment that's even more far out: the China brain. Suppose we duplicated the function of neurons in a brain, with people in a sufficiently large nation, waving flags in specific ways.

- If a stimulus causes a specific neuron in the brain to fire, the person corresponding to that neuron waves a flag with the analogous stimulus. People communicating with flags replace neurons communicating with neurotransmitters, operating precisely on the pattern of all the neurons in your brain.

- Would that nationwide complex of people with flags have thoughts? Would it be conscious? None of those seem plausible. This thought experiment seems to show that something more must be going on.

The Final Experiment

- This lecture's final thought experiment fights thought experiments on one side with a thought experiment on the other. It is Paul and Patricia Smith Churchland's response against Searle's Chinese room.

- Since the work of James Church Maxwell in 1865, we know that light and electromagnetic waves are the same thing. Now for the thought experiment: Imagine a Victorian gentleman in his parlor first reading that fact in a scientific journal. Maxwell's experiments show that light and electromagnetic waves aren't two different things. They're the same thing.

- Our gentleman isn't buying any of it. He states, "If light were a form of electromagnetism, then if a magnet were to oscillate quickly enough, the result would be visible light." He's right: That is an implication of the theory we are talking about. In a darkened Victorian parlor, he waves a magnet up and down as rapidly as he can. No light appears.

- What the Victorian gentleman is able to produce in that darkened Victorian parlor is nothing near the conditions under which magnetic oscillation

would produce visible light. The physical experiment we are imagining him performing simply isn't up to the task.

- Paul and Patricia Smith Churchland suggest that the same is true of Searle's Chinese room. We are no more able to imagine—really imagine—the complexity that would be required in order for a program to mimic a Chinese speaker than the Victorian gentleman is able to wave a magnet fast enough to produce light. The Churchlands' claim is that the clearly unacceptable conclusion in this case might not be so clearly unacceptable.

- Thought experiments or the intuitions that guide them should not be dismissed out of hand. They are a great tool for careful thinking. But they should not be a substitute for thinking.

Suggested Reading

Churchland and Churchland, "Could a Machine Think?"

Searle, "Minds, Brains, and Programs."

Questions to Consider

1. Does Mary learn something new when she leaves her black-and-white world? Does that show that there are more facts than the physical facts?

2. John Searle offers the Chinese room as an argument against "strong AI": No program, however sophisticated, could allow a computer to understand English. Paul and Patricia Smith Churchland think Searle is like the Victorian gentleman waving a magnet in a darkened room as a refutation of Maxwell's theory of electromagnetism. Where do your intuitions lie in that battle between thought experiments?

Lecture 23

Consciousness and the Explanatory Gap

This lecture faces a problem that has arisen a number of times in this course. We keep running up against the hard problem of consciousness, which has also been called the explanatory gap. The problem boils down to this: Suppose in the year 2050 we have discovered that whenever the brain displays a certain pattern and configuration, it produces consciousness. We'd still want to know why that configuration produces consciousness. What is it about that pattern that gives us subjective experience? Even when we know what produces consciousness, we'll still want to know how a physical brain function could possibly produce it.

The Hard Problem

- The hard problem of consciousness has long been with us. Credit for pressing it as a contemporary problem goes to the philosopher David Chalmers. Chalmers offers a list of things that he thinks are well within reach of our contemporary brain sciences, such as how the brain reacts to environmental stimuli and the difference between being awake and asleep.

- In each of these cases, there's a great deal we don't know. But Chalmers says all are questions of how a brain functions. We know that the brain does it. We just have to figure out how. These are the so-called easy problems—easy in the relative sense, of course.

- The really hard problem is not a question of function but of experience. Why does the brain in any particular state functioning in any particular way result in conscious experience?

Unconscious Experience

- Another way of pressing the question is to emphasize all the ways that our brains operate unconsciously, below the level of subjective experience. Take the example of a person driving to the office and suddenly realizing they have no memory of the trip: Despite avoiding pedestrians and stopping at red lights, they did this trip on "automatic pilot."

- Even the deepest and most complex thinking can occur unconsciously. The great mathematician Carl Friedrich Gauss spoke of a theorem he had tried for years to solve, without success. But then, he reported, "Like a sudden flash of lightning, the riddle was solved." He himself couldn't see what the thread was from his previous thinking to the solution that suddenly appeared before him.

- Responsiveness to the environment, intricate performance, learning, and even deep and creative formal thinking can all occur without consciousness. If the brain can do all this without consciousness, we might eventually understand how the brain performs all these functions—yet still not know how or why consciousness happens. That's the explanatory gap.

Difficulty

- Just how hard is the hard problem of consciousness? Some thinkers claim no real issue has been defined. Patricia Smith Churchland calls it the "hornswoggle problem." She points out that Chalmers' "easy" problems aren't all that easy. Even there, we often don't know what a solution is going to look like. Why think that consciousness is any different? The clear road ahead is to keep working on the neurobiology of mental phenomena in general.

- She and her husband Paul Churchland argue that our view of the issue at this point may also be limited by our current ignorance and conceptual confusion. The Churchlands predict that as scientific work proceeds we'll find that there is no explanatory gap after all.

- Daniel Dennett is another philosopher who thinks there is no real problem. While there is plenty of work to do, both philosophically and in the brain sciences, there is no sudden leap required to cross an explanatory gap. He makes a two-pronged argument.
    - Dennett argues that when we really pay attention to this thing called subjective experience, we'll find there's much less to explain than we thought.

    - Dennett also argues that a complex of different cognitive functions can explain much more than we tend to think. When we put together mechanisms for attention, for decision, for monitoring and for reporting internal states,we'll see that they give us all we need.

- Dennett outlines his two-pronged critical approach in his book *Consciousness Explained*. Criticism of Dennett accuses him of ignoring the depth of the real problem.

Activities like playing the piano appear to require great coordination and concentration, yet many people can read music essentially unconsciously.

- Another approach claims the hard problem isn't ultimately all that hard. It relies on the concept of emergence. How does subjective experience arise from a purely physical brain? Consciousness is an emergent property of a complex brain.

- The concept of emergence relies on the idea of levels. The idea is that something emerges on a higher level as a result of the interaction of elements on the lower level.

- The classical example of emergence comes from John Stuart Mill, using water. A hydrogen atom is not wet. Neither is an oxygen atom. Nor does a single molecule composed of two hydrogen atoms have that property. But put lots of those molecules together, interacting at room temperature, and you have something new: liquidity. Only now do you have something wet. That's emergence.

- The difficulty with the concept is that it relies on two requirements. One requirement is that the thing on the higher level has to be something genuinely new and different. The second requirement is that it has to be the product of what happens lower down. It has to be entirely determined by the interactions of its components.

- But if a phenomenon is determined by the action of components at a lower level, we should be able to explain it in terms of those components. We can for liquidity, in terms of atoms that slide past each other rather than locking into a lattice.

- What do we get if we propose that consciousness is an emergent phenomenon? First, we are proposing that it's genuinely new: something at one level that doesn't appear at a lower level. And second, if it is emergent, it has to be produced entirely by the operation of things at a lower level.

- How can consciousness be produced entirely by that lower level? That brings up the hard problem all over again.

Another Take

- A number of people have suggested that it's hard enough to demand radical changes in our most fundamental approaches to the world. One of those people is the distinguished physicist and mathematician Roger Penrose. There are two parts to Penrose's ideas about consciousness.

- The first part relies on important 20^{th}-century results in logic and mathematics. Penrose interprets results from the work of Alan M. Turing and Kurt Gödel as showing that mathematical insight goes beyond algorithmic systems. Our minds must thus be able to do something that algorithms and computers cannot: Consciousness must function non-algorithmically.

- The second part is a proposal for precisely how that could operate in the brain. Penrose proposes that quantum mechanics might account for non-algorithmic processing in the brain. And that it is non-algorithmic processing that is the mark of consciousness.

- There are lots of questions here. One question is whether this interpretation of Turing and Gödel's results is right. The jury's out on that one. Another question is whether there is any quantum action in the microtubules that has any effect on the brain. Many seem dubious.

Solving the Problem

- David Chalmers says it will take a radical change in our physics to solve the hard problem. He says there are fundamental aspects of our world that demand more than anything our current physics can handle. Subjective experience is the prime example.

- Consciousness should be recognized as a fundamental aspect of the universe, on par with concepts like mass and space-time. Interestingly, he adds that if we do that, our theory of consciousness will look more like a new physics than an expanded biology.

- Here again there is a range of questions. One question is what any new physics of that sort would look like or how it would operate. A second question is whether we really need a new physics: Our old physics has been astoundingly effective in both understanding and manipulating the physical world around us. A third question is whether even this would give us what we want. Suppose we rebuilt our physics by adding a new particle, the consciotron, or a new force, the C-force, and surmised that those make us conscious. Wouldn't we still have the same problem? Why should consciotrons or the C-force make us conscious?

- Chalmers is not alone, however. Galen Strawson is another philosopher who thinks the hard problem is hard enough to demand a radical change in our worldview. Strawson characterizes himself as a physicalist: Everything that exists is physical. But he also thinks that the existence of consciousness—our subjective experience—is more certain than the existence of anything else.

- Strawson argues that given those two premises, we have to recognize that consciousness is an aspect of the physical world. He doesn't take it to be merely an aspect of some small and complex corner of the physical world, contained in our skulls. For Strawson, experiential consciousness must be part of every part of the physical world.

- Revolutionary views such as those of Chalmers and Strawson, as they themselves point out, are usually met not with counterarguments but with bewildered stares: "Do you mean that rocks, ocean waves, and atoms are conscious?" To which the answer has to be "Yes," or at least, "In some sense, yes," met again with bewildered stares.

- Where else might we look for a solution to the hard problem? Dualism is always there, lurking in the background, but dualism seems to be the route that no one wants to take. The reasons are the same as those noted in earlier lectures. Divide the universe conceptually into two distinct realms, as Descartes did, and it looks like it will be conceptually impossible to put it back together.

A Hopeless Problem?

- The philosopher Colin McGinn is a mysterian. He asks whether we can solve the mind-body problem, and firmly answers no. This is a problem we cannot solve, ever and in principle.

- It's not that there isn't an answer. The universe knows the answer. But how the physical brain creates conscious experience is forever and inevitably closed to us. McGinn's analogy is the blind spot. There are things that are there to be seen, but you cannot see them because they are in your blind spot. Somehow the physical does meet the mental in the operation of the universe, but that meeting point lies where we could never see it: It lies in our intellectual or conceptual blind spot.

- Is it possible that the answer we seek is one that we can't in principle find? The answer may be yes. There are scientific principles that are beyond the conceptual reach of raccoons. There are aspects of mathematics that are beyond the reach of chimpanzees. There seems no reason to think that there aren't also things that are beyond our reach.

- On the other hand, there are things that might be true but that it would be intellectually wrong to accept. Why? Because that acceptance would block the possibility of intellectual progress by shutting down further inquiry.

- If we decide there is no answer, we will no longer look for one. If we don't look for one, we won't find one. That could be an intellectual loss beyond measure. Suppose there really is an answer, lingering just beyond reach.

Suggested Reading

Chalmers, "Facing Up to the Problem of Consciousness."

Nagel, "What Is It Like to Be a Bat?"

Penrose, *The Emperor's New Mind*.

Questions to Consider

1. This lecture lists a number of things that we seem to be able to do without consciousness. Is there something that we cannot do without consciousness?

2. How hard do you think the hard problem of consciousness is?

Lecture 24

A Philosophical Science of Consciousness?

From the start of this course, we've been grappling with the mind-body problem. Much of the material we've covered suggests the need for both better philosophy of consciousness and a better science of mind. This lecture suggests that we also need an integration of the two: a philosophical science of consciousness. How might we get there? What might a philosophical science of consciousness look like? Those are the questions this lecture tackles.

The Scientific Approach

- In additional to the philosophical approach and the scientific approach, we've also considered the information-theoretic approach. Here the idea is that we can learn about the mind by trying to build information-processing devices that mimic mental functioning.

- In which of these—science, philosophy, or information—lies the answer to understanding mind, body, and consciousness? Taken alone, none of these can give us the answers we're after. The hard problem of consciousness illustrates that.

- Those working in the brain sciences are generally careful to specify what they're doing as a search for "neural correlates of consciousness." They hope for

a map of brain dynamics that correlates with states of consciousness. The ideal outcome would be an identifiable brain process—let's call it brain process C—that is precisely what a conscious brain is doing.

- We would be able to say, "When C occurs in a brain, you have a conscious brain. When C does not occur, you don't."

- We are very far from that ideal. We don't know the neural correlates of consciousness. We can't identify any such brain process C.

- Should we keep looking? Absolutely. The search for C is a search for an important part of what we want to know about consciousness and the mind. We need to know what is happening in the brain. Nevertheless, even full achievement of that scientific goal wouldn't tell us everything that we want to know.

- Sometimes the problem is phrased in very strong terms. The realm of science is the realm of objectivity. But the realm of consciousness is precisely the opposite. It is the realm of the subjective, the essentially private. The idea of a science of consciousness is therefore a contradiction in terms, or so the reasoning goes.

- There is a way of taking that line of reasoning that seems correct, but also a way that might lead us astray. It might lead us astray if we think that we could never do scientific experiments on neural correlates of consciousness. A correlate would link something objective about the brain with an aspect of consciousness or subjective experience. If we refuse an objective approach to the subjective, we'd never be able to include that subjective side in our scientific investigations.

- Experiments regarding consciousness have played a major role throughout this course: Recall V. S. Ramachandran's experiments on the sensations felt in phantom limbs. We've reviewed Ronald Myers, Roger Sperry, and Michael Gazzaniga's experiments on experience and sense of self in split-brain patients. We've looked at Grey Walter and Benjamin Libet's experiments on whether the body is already in action before a conscious decision is made, and so on.

- It's true that in any scientific experiment regarding consciousness, we have to rely on first-person reports of consciousness. We know that those can go wrong. We have seen that we can be wrong about characteristics of our own consciousness. But there is no reason to doubt there is a genuine phenomenon being reported.

- The real conclusion is simply that objective access to the subjective realm is necessarily indirect. That's not so different from much of the rest of our science, which similarly relies on indirect objective indicators. For instance, our knowledge of the distance and velocity of other galaxies is based on inferences regarding the shift in the color of light that we receive from those galaxies.

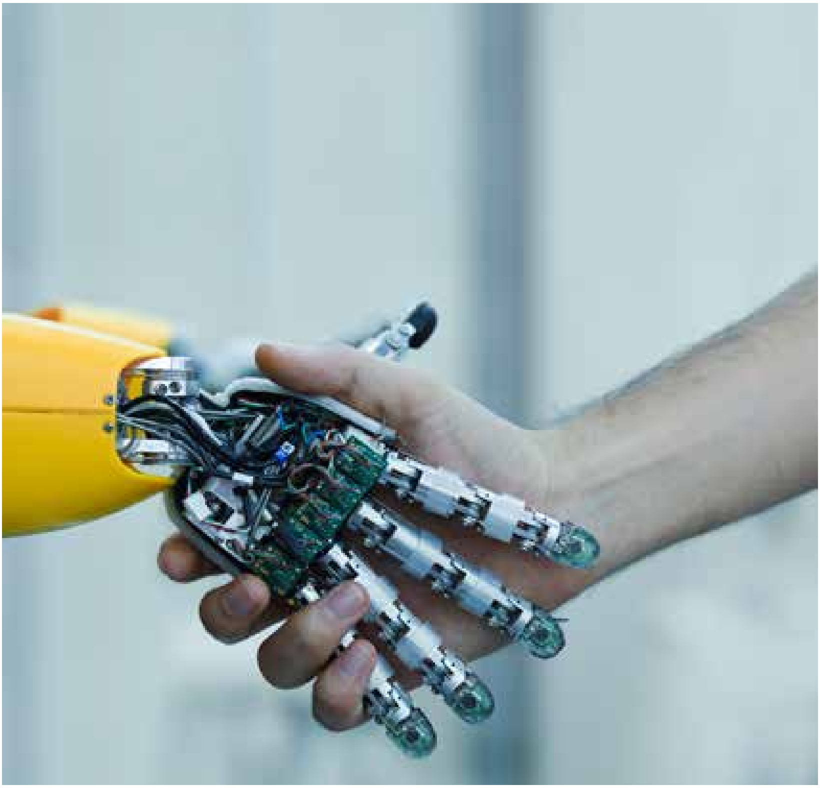
The Information-Theoretic Approach

- Science is taking us farther each day in the search for neural correlates of consciousness. But the understanding we're after asks for more than correlates. We'll need more than science alone—or at least more than that kind of science.

- Let's consider an information-theoretic approach. From all appearances, the brain is a massive parallel information-processing device. If we can understand how that information processing occurs, perhaps we can understand more about intelligence, cognition, and perception. Maybe we can even understand more about consciousness.

- But the history of artificial intelligence is a history that aimed for computational intelligence, not computational consciousness. The computer Deep Blue plays world-class chess. IBM's Watson beat human players on *Jeopardy!* But no one thinks Deep Blue or Watson is conscious.

- The history of artificial intelligence is a checkered one. It begins with the Turing Test, a test that is entirely behavioral. Turing thought that by 2000 we'd have a machine that with a probability of 70 percent would be indistinguishable from a human after five minutes of questioning. We haven't passed that benchmark yet.



- But when it comes to artificial intelligence, we continue to see areas of progress and glimmers of hope. A major sign of progress is the shift from linear programming to structures that imitate human neurons. Machine learning constitutes a prolific and growing field, in large part because it can be used to discern patterns in the torrents of data now available to us.

- Marvin Minsky, a major figure in the history of artificial intelligence, said that "mind is what the brain does." There certainly is something the brain does. In principle, we should be able to specify what that something is.

- But in practice, the whole path of investigation turns out to be a lot harder. There are many more blind alleys than we ever thought, and we've learned that our minds don't function like computers.

Philosophy

- Neither the brain sciences alone nor informational sciences alone seem capable of giving us a real understanding of consciousness. Should the question then be left to philosophers?

- This isn't a good idea. Part of the argument is inductive. We've traced the history of philosophy of mind over more than 2,500 years. Philosophical work alone has taken us no further than we are now. Philosophers are great at posing questions, but not so great at agreeing on answers.

- The philosopher's strengths are logical, conceptual, and analytic. Those are precisely what we need in order to figure out what the central questions are. They are precisely what we need in order to get a handle on how one might be able to answer those questions. But to go to a philosopher for a final answer on a question is to make a mistake about philosophy's role.

- Philosophers are going to be of most use when their logical methods and conceptual talents are applied to concepts in use—concepts of interest and importance for a science of mind and consciousness, for example.

Loops

- Taken alone, none of the disciplines we've looked at—philosophy, artificial intelligence, or brain sciences—seem capable of giving us the understanding of mind, body, and consciousness that we're after. Could they do better working together? It's tough to definitely answer yes, but integration of the three offers progress in a way that none of the three does alone.

- Many thinkers agree that loops are an important key to consciousness. There is something about consciousness that is essentially self-referring or reflexive. Consciousness is somehow a mental state that loops on itself—that includes itself in its field of attention.

- We are creatures that interact with the environment using "fast and frugal" environmental detectors. We have a brain built to interact with our

environment in terms of fast and frugal detectors for color, sound, motion, temperature and touch.

- In order to respond in complex ways to a complex environment, our brains prioritize and categorize that fast and frugal input. But it may be important when reacting to something new, or preparing for the unexpected, to attend to the processing categories themselves: categories of color, of sound, of movement. The value of consciousness may lie in the availability of that kind of loop: an ability to monitor and revise categories of interaction with the environment in the process of interaction and on the fly.

- Regarding the brain: Processing doesn't proceed linearly from one area to another in sequence. There is always backflow to the earlier areas. A theory of cognitive loops might tell us why.

A Plan

- How might we work toward a philosophical science of mind, body, and consciousness? The first step is to figure out what consciousness is by figuring out what consciousness is for: What does it do that other cognitive processing could not?

- The second step is to analyze the process in abstract terms. From input to output, what function is needed to produce the process we've identified as what consciousness is for?

- Once we have an abstract characterization of the process, we need to move to the concrete specifics of the brain: How does the brain perform that function?

- None of those steps will be easy, and they present some paradoxical possibilities. For instance, regarding the first step, identifying what consciousness is for: We are evolved creatures, shaped by natural selection, but not everything about us was specifically selected for. Our blood is red, but nature didn't care about the color of our blood. Likewise, it is possible consciousness just came along for the ride—a cosmic accident.

- That aside, suppose there is something that consciousness does, and that we can put our finger on what that is. The next step would be to specify that functional something operationally. It must be a process that moves from some range of inputs to some range of outputs.

- Can we give a formal description of that process? Those working in artificial intelligence, computer science, and aspects of mathematics are important at this point in the project.

- There is a second paradox buried in this step. Suppose we succeed in giving a formal outline of the process of consciousness. There shouldn't then be any obstacle to building that formal process into a computer. The paradoxical possibility is that if we succeed at this stage, we will have all we need in order to build conscious machines.

- The reasoning that leads to that second paradox is not as tight as the first. It only holds if the process outlined for consciousness is a machine-like process: an algorithm in the sense of Turing and Penrose.

- There is also another possibility. It might turn out that the process we specify abstractly can't be built concretely into the kinds of machines we have on hand. It might require a very different kind of machine. It might require biological wetware.

- It is at the third stage that we need the brain sciences. If we can establish what consciousness does, and if we can specify in the abstract the process required to do that, we will want to know the concrete form that process takes in the brain. What parts of the brain do the work of which parts of the formal process? And precisely how do they do it?

The Link

- Were we able to complete each step in such a research program, we would be far closer to the kinds of answers we've been looking for. What is the link between the material brain and the world of subjective experience?

- The answer would go something like this: In the first step, we would be able to see that consciousness has been evolutionarily selected for because it does whatever it does. In the second step, the formal sciences would give us a functional description of the process involved. In the third step, work in the brain sciences would show us how the brain makes that abstract function concrete. Might that not give us an answer to the hard problem?

- It is possible that the hypothesis that we'll never know how consciousness happens is correct. But if accepted, it would guarantee that we'll never know. We have an intellectual obligation to try to figure things out—including consciousness.

- If progress is possible, it will demand some integration of disciplines. We will certainly need the hard data of the brain sciences. We may well need the tools of the formal sciences. But neither of those will be enough. We will need conceptual philosophical work in order to outline what consciousness really is.

Suggested Reading

Baumgartner and Payr, *Speaking Minds*.

Blackmore, *Conversations on Consciousness*.

Grim, ed., *Mind and Consciousness: 5 Questions*.

Questions to Consider

1 This concluding lecture looks back on the resources we've drawn on from the brain sciences, from information theory, and from philosophy. When you think back on the lectures, which are the contributions you remember most?

2 This lecture argues that a philosophical science of consciousness will require an integration of work from all three fields. In which area do you think we should put the most work toward that integration?

Bibliography

Afriat, Alexander and Franco Selleri. *The Einstein, Podolsky and Rosen Paradox in Atomic, Nuclear, and Particle Physics*. New York: Plenum Press, 1999. A nicely complete guide to the EPR thought experiment and its conceptual fallout.

Aquinas, Thomas. Tr. James H. Robb. *Questions on the Soul*. Milwaukee, WI: Marquette University Press, 1984. Circa 1265. Aquinas's attempt to merge Aristotle with Christianity on the soul.

Aristotle. *De Anima*. Clarendon Aristotle Series. Tr. Christopher Shields. Oxford: Oxford University Press, 2016. Circa 350 B.C. Aristotle's functionalism, highlighted in Lecture 2.

Augustine. *The Greatness of the Soul, The Teacher*. Tr. Joseph M. Colleran. Eds. Johannes Quasten and Joseph Plumpe. Circa 387 AD. New York: the Newman Press, 1950. Augustine's Platonic view of the soul.

———. *City of God*. Tr. Marcus Dods. New York: Modern Library, 1993. Available in many editions. Circa 415 A.D. Augustine's masterpiece, including an anticipation of Descartes's argument.

Baars, Bernard. *In the Theater of Consciousness: The Workspace of the Mind*. Oxford: Oxford University Press, 1997. Baars's most complete outline of his global workspace theory of consciousness.

Baars, Bernard, William P. Banks, and James B. Newman, eds. *Essential Sources in the Scientific Study of Consciousness*. Cambridge, MA: MIT Press, 2003. A solid collection of reprinted pieces, arranged by aspects and functions of consciousness.

Baumgartner, Peter and Sabine Payr. *Speaking Minds: Interviews with Twenty Eminent Cognitive Scientists*. Princeton, NJ: Princeton University Press, 1995. One of several collections of anthologies with contemporary thinkers, always revealing of the full human personalities behind the theories.

Beakley, Brian and Peter Ludlow, eds. *The Philosophy of Mind: Classical Problems/Contemporary Issues*. Cambridge, MA: MIT Press, 2006. A compilation of pieces attempting to link pieces in the history of philosophy with contemporary debates.

Bear, Mark F., Barry W. Connors, and Michael A. Paradiso, eds. *Neuroscience: Exploring the Brain*. 2nd edition. Baltimore, MD: Lippincott Williams & Wilkins, 2001. A very thorough textbook.

Beilock, Sian. *How the Body Knows Its Mind: The Surprising Power of the Physical Environment to Influence How You Think and Feel*. New York: Atria Books, 2015. Occasionally breathless with an air of self-help, Beilock's book is nonetheless one of the best available overviews of research on the impact of body on mind.

Beowulf. Tr. Burton Raffel. New York: Signet, 2008. The oldest surviving epic in Old English. Available in many versions.

Blackmore, Susan. *Consciousness: An Introduction*. Oxford: Oxford University Press, 2004. Written and illustrated like an undergraduate textbook, but drawing on a wealth of background research and with an enjoyable Blackmore spin.

———. *Consciousness: A Very Short Introduction*. Oxford: Oxford University Press, 2005. Short but wonderfully compact. Points touched on just briefly here are expanded in Blackmore's *Consciousness: An Introduction*.

———. *Conversations on Consciousness*. Oxford: Oxford University Press, 2006. One of several collections of anthologies with contemporary thinkers, always revealing of the full human personalities behind the theories.

Block, Ned. "Troubles with Functionalism." *Minnesota Studies in the Philosophy of Science* 9 (1978): 261–325. Block's Chinese brain thought experiment, with other critiques of functionalism.

Block, Ned, Owen Flanagan, and Güven Güzeldere, eds. *The Nature of Consciousness: Philosophical Debates*. Cambridge, MA: MIT Press, 1997. A solid collection of reprinted pieces, arranged by philosophical issues.

Bloom, Deborah. *Ghost Hunters: William James and the Search for Scientific Proof of Life After Death*. New York: Penguin, 2007. Briefly mentioned in the lectures, James's involvement in psychical research proved both fascinating and frustrating. Bloom writes the history like a novel.

Borges, Jorge Luis. "Funes the Memorious." In Anthony Kerrigan, ed., *Ficciones*, 107–116. New York: Grove Press, 1962. A wonderful tale of what Locke forgot in his theory of memory: forgetting.

Bresloff, P. C., J. D. Cowan, M. Golubitskyh, P. J. Thomas, and M. C. Wiener. "What Geometric Visual Hallucinations Tell Us About the Visual Cortex." *Neural Computation* 14, no. 2 (2002): 473–491. Cowan and Bresloff's mathematical modeling of standard hallucination geometrics as an effect of the structural organization of the visual cortex. Dense but fascinating.

Brooks, Rodney. "Intelligence without representation." *Artificial Intelligence* 47, nos. 1–3 (1991): 139–159. Inventive ideas, conversationally expressed, from a primary theoretician in artificial intelligence and practical master of robotics.

Cannon, W. B. "The James-Lange Theory of Emotions: A Critical Examination and an Alternative Theory. *The American Journal of Psychology* 39 (1927): 106–124 and *The American Journal of Psychology* 100, Special Centennial Issue, nos. 3–4 (1987): 567–586. The original source for the Cannon-Bard theory of emotion.

Carter, Rita. *Mapping the Mind*. Berkeley: University of California Press, 1999. Despite a text that sometimes seems scattered, a wonderful introduction to the brain sciences, prolifically illustrated.

Casti, John. *The Cambridge Quintet*. Reading, MA: Addison-Wesley, 1998. A series of fictional dinner conversations on minds and machines between Alan M. Turing, Ludwig Wittgenstein, Erwin Schrödinger, and J. B. S. Haldane, and C. P. Snow.

Chalmers, David. "Facing Up to the Problem of Consciousness," *Journal of Consciousness Studies* 2, no. 3 (1995): 200–219. Chalmers's influential formulation of the hard problem of consciousness.

———. *The Conscious Mind: In Search of a Fundamental Theory*. New York: Oxford University Press, 1996. Chalmer's most complete development of panpsychism as an answer to the hard problem of consciousness.

Churchland, Patricia Smith. *Brain-Wise: Studies in Neurophilosophy*. Cambridge, MA: MIT Press, 2002. Reports and reflections from a philosopher immersed in the details of neuroscience.

———. *Neurophilosophy: Toward a Unified Science of the Mind/Brain*. Cambridge, MA: MIT Press, 1986. Rich in neurophysiological detail, the book also lays out P. S. Churchland's most fully developed defense of reductionism.

———. "The Hornswoggle Problem." *Journal of Consciousness Studies* 3, nos. 5–6 (1996): 402–408. Churchland's argument that we have no right to assume that the hard problem of consciousness is any harder than any of the many other questions of the mind for which we don't currently have answers.

———. *The Engine of Reason, the Seat of the Soul*. Cambridge, MA: MIT Press, 1995. Particularly valuable for an approach to questions of the mind with the artificial intelligence model of recurrent neural networks.

Churchland, Paul and Patricia Smith Churchland. "Could a Machine Think?" *Scientific American* 262, no. 1 (January 1990): 32–37. The Churchlands' response to Searle's Chinese room thought experiment.

Clark, Andy and David Chalmers. "The Extended Mind." *Analysis* 58, no. 1 (1988): 7–19. Reprinted in Patrick Grim, Kenneth Baynes, Peter Ludow and Gary

Mar, eds., *The Philosopher's Annual*, vol. 21, 59–74. Atascadero, CA: Ridgeview 2000. Clark and Chalmers's presentation of Otto and Inge, arguing for a mind that extends into the world beyond skull and skin.

Cohen, Jonathan D. and Jonathan W. Schooler, eds. *Scientific Approaches to Consciousness*. Mahway, NJ: Lawrence Erlbaum, 1997. A collection of often fairly technical pieces on specific topics.

Copleston, Frederick. *A History of Philosophy*. Nine volumes. New York: Image Books, 1994. Despite the series's history, originally published in 1960 as a history of philosophy for Catholic seminary students, Copleston's remains an important secondary source across the history of philosophy as a whole.

Coterill, Rodney. *Enchanted Looms: Conscious Networks in Brains and Computers*. Cambridge: Cambridge, University Press, 1998. Sometimes wandering, always fascinating. Includes a thorough discussion of Benjamin Libet and Roger Carpenter on consciousness and timing.

Crick, Francis. *The Astonishing Hypothesis*. New York: Touchstone, 1994. A wide-ranging guide to the brain, including but by no means limited to Crick's 40-hertz hypothesis for consciousness. Crick's passion for finding out the real answers to mental functioning is obvious throughout.

Crick, Francis and Christof Koch. "What Is the Function of the Claustrum?" *Philosophical Transactions of the Royal Society B* 360, issue 1458 (June 2005): 1271–1279. Crick's second hypothesis regarding consciousness, a piece he was working on when he died.

Damasio, Antonio. *Descartes' Error: Emotion, Reason, and the Human Brain*. New York: Penguin, 1994. Damasio says that Descartes's error was to leave out emotion. A developed theory of a body-based somatic marker hypothesis.

Damasio, Antonio. *The Feeling of What Happens: Body and Emotion in the Making of Consciousness*. Orlando, FL: Harcourt 1999. Damasio sees his theory of emotion as a descendant of the James-Lange theory and takes pains to defend it against later critiques.

Darwin, Charles. *The Expression of the Emotions in Man and Animals*. Anniversary edition. Oxford: Oxford University Press, 2009. First published 1872. An admirable followup to *The Origin of Species* and *The Descent of Man*, taking implications for the evolution of human psychology head on.

Dehaene, Stanislas. *Consciousness and the Brain*. New York: Penguin, 2014. A full-volume review of research on consciousness, with a clear exposition of the author's own fascinating contributions.

Dennett, Daniel. *Brainstorms: Philosophical Essays on Mind and Psychology*. Cambridge, MA: MIT Press, 1985. A collection of essays presented at conferences and departmental colloquia, collected as one of Dennett's earliest and best volumes.

———. *Consciousness Explained*. New York: Penguin, 1991. Sometimes mocked as "Consciousness Explained Away," Dennett's full critical take on qualia and the Cartesian theater with his multiple drafts model offered as a positive alternative.

———. "Quining Qualia." In A. J. Marcel and E. Bisiach, eds., *Consciousness in Contemporary Science*, 43–77. Reprinted in Ned Block, Owen Flanagan and Güven Güzeldere, eds., *The Nature of Consciousness: Philosophical Debates*, 619–642. Cambridge, MA: MIT Press, 1997. Dennett's thought experiments against the "raw feels" of subjective qualia.

———. "The Unimagined Preposterousness of Zombies." *Journal of Consciousness Studies* 2, No. 4 (1995): 322–326. Dennett's attack on Flanagan and Polger's zombie thought experiments.

———. "Where Am I?" In Dennett, *Brainstorms: Philosophical Essays on Mind and Psychology*, 310–323. Cambridge, MA: MIT Press, 1985. Also in Douglas Hofstadter and Daniel Dennett, eds. *The Mind's I*, 217–229. New York: Basic Books, 2000. A wonderful piece of fictional autobiography exploring the question of personal identity.

Dennett, Daniel and Marcel Kinsbourne. "Time and the Observer: The Where and When of Consciousness in the Brain," *Behavioral and Brain Sciences* 15, no. 2 (1992): 183–244. Reprinted in Patrick Grim, Gary Mar, and Peter Williams eds.,

The Philosopher's Annual, vol. 21 (1992), Ridgeview Press, 23–68. A consideration of Libet's and other research, arguing for a multiple drafts view on which there is no "finish line" for consciousness in the brain.

Descartes, René. *Discourse on Method and Meditations on First Philosophy*. Tr. Donald A. Cress. Hackett Publishing, 1999. First published in 1637 and 1641. *The core Descartes*. Also available in many other versions and collections.

Desikachar, T. K. V. *The Heart of Yoga*. Rochester, VT: Inner Traditions, 1995. A valuable and accessible introduction to the conceptual core of yoga.

Diogenes Laertes. *Lives of Eminent Philosophers*, vols. 1 and 2. Cambridge, MA: Harvard University Press, Loeb Classical Library, 1972. Circa 250 B.C. The classic source on ancient Greek philosophers, in Greek and English format.

Edelman, Gerald M. *Bright Air, Brilliant Fire: On the Matter of the Mind*. New York: Basic Books, 1992. A major source for Edelman's neural Darwinism, but valuable in other regards as a set of reflections worth rereading.

Flanagan, Owen. *Dreaming Souls: Sleep, Dreams, and the Evolution of the Conscious Mind*. Oxford: Oxford University Press, 2000. Flanagan writes some of the most scientifically informed philosophy of mind available today, and is known in particular for his work on sleep and dreams. As outlined in the lectures, Flanagan argues that dreams are spandrels: unselected side effects of the evolutionary process.

———. *The Science of Mind*. 2nd edition. Cambridge, MA: MIT Press, 1984. An ambitious attempt at a contemporary philosophical overview of major movements in psychology.

Flanagan, Owen and Thomas Polger, "Zombies and the Function of Consciousness." *Journal of Consciousness Studies* 2, no. 4 (1995): 313–321. A defense of imaginable philosophical zombies as a negative indicator for functionalism. Includes the zombie scorecard.

Foer, Joshua. *Moonwalking with Einstein: The Art and Science of Remembering Everything*. New York: Penguin, 2012. An entertaining personal foray into the world of competitive mnemonics.

Freud, Sigmund. "Project for a Scientific Psychoanalysis." 1895. Available at http://users.clas.ufl.edu/burt/freud%20fleiss%20letters/200711781-013.pdf. Freud's earliest outline for a science of psychology.

Galileo. *The Essential Galileo*. Ed. and trans. Maurice A. Finocchiaro. Indianapolis, IN: Hackett, 2008. Circa 1610 to 1632. A good collection, including Galileo on Aristotle.

Gazzaniga, Michael. "The Split Brain in Man." *Scientific American* 217, no. 2 (1967): 24–29. Gazzaniga's later experiments and reflections on strange cases of consciousness in the split brain.

Genova, Lisa. *Left Neglected*. New York: Gallery Books, 2011. A compelling novel of parietal neglect, in which the left side the world disappears.

Gibson, James J. *The Ecological Approach to Visual Perception*. Hillsdale, NJ: Lawrence Erlbaum 1986. Rich in detailed analysis of perception, the core of Gibson's work remains his theory of affordances.

———. "The Theory of Affordances." In R. Shaw and J. Bransford, eds., *Perceiving, Acting, and Knowing: Toward an Ecological Psychology*. Hillsdale, NJ: Lawrence Erlbaum, 1977, pp. 67–82. The core of Gibson's affordance theory, accessibly presented and without excessive detail.

Goetz, Stewart and Charles Taliaferro. *A Brief History of the Soul*. Chichester, West Sussex, UK: Wiley-Blackwell, 2011. Combines a review of the soul in the history of philosophy with a less successful attempt at a contemporary outline and defense.

Grim, Patrick, ed. *Mind and Consciousness: 5 Questions*. Automatic Press/VIP, 2009. One of several collections of anthologies with contemporary thinkers, always revealing of the full human personalities behind the theories.

Grim, Patrick and Nicholas Rescher. *Reflexivity: From Paradox to Consciousness*. Ontos-Verlag 2012. An attempt to carry the theme of loops from semantic paradox through issues of indexicals to questions of the structure of consciousness.

Gutenplan, Samuel, ed. *A Companion to the Philosophy of Mind*. Oxford: Blackwell, 1995. An encyclopedia format with entries on both topics and authors. Valuable for Gutenplan's extensive introduction alone.

Hebb, Donald. "On Watching Myself Get Old." *Psychology Today* 12, no. 6 (1978): 15–23. Autobiographical observations on personal aging by a thinker of wide influence in both psychology and computational modeling.

Hobbes, Thomas. *Leviathan*. Norton Critical Edition. Eds. Richard E. Flathman and David Johnston. New York: W. W. Norton, 1997. First published in 1651. Hobbes's major work, targeted to social and political philosophy but grounded in a thorough-going materialism. Also available in many other versions.

Hoffmann, Albert. *LSD, My Problem Child: Reflections on Sacred Drugs, Mysticism and Science*. Sarasota, FL: Multidisciplinary Association for Psychedelic Studies, 2009. Reflections on drug effects and their context from the discoverer of LSD.

Hoffman, Donald D. *Visual Intelligence: How We Create What We See*. New York: W. W. Norton, 1998. A wonderfully inventive book exploring and proposing psychological theory for visual illusion.

Hofstadter, Douglas and Daniel Dennett, eds. *The Mind's I*. New York: Basic Books, 2000. A collection of pieces in philosophy of mind with an emphasis on fiction, offering a light-hearted collection that raises hard-headed issues.

Homer. *Iliad*. Tr. Robert Fagles. New York: Penguin, 1998. Circa 760 BC. Valuable for innumerable reasons, including the early Greek view of the soul. Available in many editions.

———. *Odyssey*. Tr. Robert Eagles. New York: Penguin, 1998. Circa 760 BC. Valuable for innumerable reasons, including the early Greek view of the soul. Available in many editions.

Hume, David. *A Treatise of Human Nature*. Ed. David Fate Norton. Oxford: Oxford University Press, 2001. Originally published 1738. Available in many editions, and free online. The earlier and in some ways fresher of Hume's two major works on mind, a landmark in the history of philosophy.

Hume, David. *An Enquiry Concerning Human Understanding*. Ed. with an introduction by Peter Millican. Oxford: Oxford University Press, 2008. Originally published 1748. Available in many editions, and free online. The other and later of Hume's two major works on mind.

———. *Dialogues Concerning Natural Religion*. Ed. Richard H. Popkin. Indianapolis, IN: Hackett, 1980. First published 1779. Available in many editions. Includes Hume's canonic treatment of the problem of evil.

Jackson, Frank. "What Mary Didn't Know." *Journal of Philosophy* 83, no. 5 (1986): 291–295. Not the first appearance of Jackson's black-and-white Mary, but the clearest. Mary appeared earlier in "Epiphenomenal Qualia," *Philosophical Quarterly* 32 (1982), 127–136.

James, William. "The Dilemma of Determinism." In *The Will to Believe*, 105–130. Cranston, RI: Angelnook, 2012. Also available in other editions and collections. James's analysis of the classical problem of free will and determinism.

———. "On Some Hegelisms." *Mind* 7, no. 26 (1882): 186–188. Excerpted as "Subjective Effects of Nitrous Oxide" in Patrick Grim, *Philosophy of Science and the Occult*, 2nd edition, 355–360. Albany, NY: SUNY Press, 1990. James's jocular but characteristically reflective report on his own experiences under nitrous oxide.

———. *The Principles of Psychology*. Two volumes. New York: Dover Publications, 1950. Originally published in 1890 by Henry Holt and Company. This is always instructive and often inspiring to dip into.

———. *Psychology: The Briefer Course*. New York: Dover Publications, 1950. Originally published in 1892 by Henry Holt and Company. This is James's abridgment of the longer work, with chapters covering stream of consciousness, the self, attention, memory, imagination, reasoning, and emotion.

———. "What Is An Emotion?" *Mind* 9, no. 34 (1884): 188–205. Available at http://psychclassics.yorku.ca/James/emotion.htm. James's classic piece on emotion.

James, William. *William James on Psychical Research*. Eds. Garner Murphey and Robert O. Ballou. New York: Viking Press, 1960. Briefly mentioned in the lectures, James's involvement in psychical research proved both fascinating and frustrating. This volume collects his writings on the topic.

Kandel, Eric. *In Search of Memory: The Emergence of a New Science of Mind*. New York: W. W. Norton, 2006. A Nobel Prize winner's charmingly autobiographical tour through recent discoveries in memories and memory formation.

Keenan, Julian, Gordon Gallup, and Dean Falk. *The Face in the Mirror: How We Know Who We Are. N*ew York: HarperCollins 2003. An accessible and entertaining review of the Gallup mirror test and related work on self-consciousness.

Kirk, G. S. and J. E. Raven. *The Presocratic Philosophers: A Critical History with a Selection of Texts*. New York: Cambridge University Press, 1984. The canonical collection of Presocratic fragments.

Kurzweil, Ray. *The Singularity is Near: When Humans Transcend Biology*. New York: Penguin, 2006. Kurzweil's futuristic predictions.

LeDoux, Joseph. *The Emotional Brain: The Mysterious Underpinnings of Emotional Life*. New York: Simon & Schuster 2015. A major source for the contemporary view outlined in the lecture on emotions.

———. *The Synaptic Self: How Our Brains Become Who We Are*. New York: Penguin, 2002. With reflections on the concept of the self, the core of the book is LeDoux's important work on emotions.

Leibniz, Gottfried Wilhelm. *Discourse on Metaphysics and Other Essays*. Tr. Robert Arlew and Daniel Garber. Indianapolis, IN: Hackett, 1991. Leibniz's arguments against mechanism and in defense of dualism in the Monadology (1714) and other sources. Pieces available in many editions.

Levine, Marvin. *The Positive Psychology of Buddhism and Yoga*. Hove, East Sussex: Routledge, 2009. A very personal and practical take on traditions from the East by an eminent psychologist. Deep insights wrapped in a common touch.

Libet, Benjamin. "Unconscious Cerebral Initiative and the Role of Conscious Will in Voluntary Action." *Behavioral and Brain Sciences* 8, no. 4 (1985): 529–566. Libet's research indicating that conscious decision comes too late to be an act of initiating free will.

Locke, John. *An Essay Concerning Human Understanding*. Ed. Roger Woolhouse. New York: Penguin Classics, 1998. First published in 1690. A canonical text of British empiricism, with Locke's theory of memory-bound personal identity. Available in many editions.

Loftus, Elizabeth, and Katherine Ketcham. *The Myth of Repressed Memory: False Memories and Allegations of Sexual Abuse*. New York: St. Martin's Griffin, 1994. A psychologically deep examination of a wrenching issue by a world expert on memory and eyewitness testimony.

Lucretius. *The Nature of Things*. Tr. Alicia Stallings. New York: Penguin, 2007. Circa 50 B.C. The Epicurean take on determinism and free will, including atoms that occasionally "swerve."

Malebranche, Nicolas. *Philosophical Selections*. Ed. Steven Nadler. Indianapolis, IN: Hackett, 1991. Circa 1675. Malebranche's occasionalist defense of dualism.

Mashour, George. "Integrating the Science of Consciousness and Anesthesia." *Anesthesia and Analgesia* 103, no. 4 (2006): 975–82. An accessible but elegant argument that studies in anesthesia offer clues to understanding consciousness.

Mashour, George and Michael Alkire. "Evolution of Consciousness: Phylogeny, ontogeny, and emergence from general anesthesia." *Proceedings of the National Academy of Sciences USA* 110, Supplement 2 (June 18, 2013): 10,357–10,364. Wider reflections on the significance of studies in anesthesia for an understanding of consciousness.

McGinn, Colin. "Can We Solve the Mind-Body Problem?" *Mind* 98, no. 391 (1989): 349–366. McGinn is a mysterian, arguing that the answer is no.

Miller, George. "The Magical Number Seven, Plus or Minus Two: Some Limits on Our Capacity for Processing Information." *Psychological Review* 63, no. 2 (1956): 81–97. One of the most cited papers in psychology, a touchstone for the limitations of short-term or working memory.

Minsky, Marvin. *The Society of Mind*. New York: Touchstone, 1985. The overall aim of the book is a theory of intelligent minds built from nonintelligent agents. Minsky's one-page sections harbor a wealth of inventive ideas.

Minsky, Marvin and Seymour Papert. *Perceptrons, Expanded Edition*. Cambridge, MA: MIT Press, 1987. Minsky and Papert's devastating critique of Frank Rosenblatt's groundbreaking neural networks. A quotation from the introduction is used in Lecture 16.

Miranda, Robbin A., William D. Casebeer, Amy M. Hein, et al. "Darpa-funded efforts in the development of novel brain-computer interface technologies." *Journal of Neuroscience Methods* 244 (April 2015): 52–67. An overview of a number of the computer-brain interface technologies mentioned in Lecture 1.

Moravec, Hans. *Mind Children: The Future of Robot and Human Intelligence*. Cambridge, MA: Harvard University Press, 1990. Moravec predicts that our real descendants will be robotic rather than biological.

Nagel, Thomas. "Brain Bisection and the Unity of Consciousness." *Synthese* 22, nos. 3–4 (1971): 396–413. Nagel's attempt to make philosophical sense of split-brain cases.

———. *Mind & Cosmos*. Oxford: Oxford University Press, 2012. Nagel's extended and developed attack on "neo-Darwinian" materialism.

———. *The View From Nowhere*. Oxford: Oxford University Press, 1986. Classic Nagel, examining aspects of, in his words, "a single problem: how to combine the perspective of a particular person inside the world with an objective view of that same world, the person and his viewpoint included."

———. “What Is It Like to Be a Bat?” *Philosophical Review* 83, no. 4 (1974): 435–350. Reprinted in Brian Beakley and Peter Ludlow, *The Philosophy of Mind*, 255–266. Cambridge, MA: MIT Press 2006. And in Douglas Hofstadter and Daniel Dennett, *The Mind’s I*, 391–402. New York: Basic Books, 2000. Nagel’s groundbreaking piece, historically important for refocusing philosophical attention on the problem of subjective experience.

Nauriyal, D. K., M. S. Drummond, & Y. B. Lal. *Buddhist Thought and Applied Psychological Research*. New York: Routledge, 2006. A nice attempt to interface points in Buddhism with contemporary work in psychology.

Neurath, Otto and Rudolf Carnap, eds. *International Encyclopedia of Unified Science*, vol. 1, nos. 1–5. University of Chicago Press, 1955. The high-water mark of materialistic reductionism.

Nietzsche, Friedrich. *Philosophy During the Tragic Age of the Greeks*. Tr. Marianne Cowan. Washington DC: Regnery, 1962. An incomplete work circa 1873. Nietzsche’s reflections on the beginning of Greek philosophy in Thales’s “preposterous fancy” are used in Lecture 2.

O’Connor, D. J., ed. *A Critical History of Western Philosophy*. New York: Free Press, 1964. A chronologically arranged set of essays on major figures in the history of philosophy, both expository and critical.

Parfit, Derek. *Reasons and Persons*. Oxford: Clarendon Press, 1984. An enormously impressive piece of sustained conceptual work with a primary emphasis on ethics, but not ethics alone.

———. “The Unimportance of Identity.” In H. Harris, ed., *Identity*, 3–88. Oxford: Oxford University Press, 1995. Parfit’s transporter example, with his reflections on what it shows us about what we should care about.

Penrose, Roger. *The Emperor’s New Mind: Concerning Computers, Minds, and the Laws of Physics*. New York: Oxford University Press, 1989. Penrose’s argument that mathematical insight goes beyond the limits of algorithms.

———. *Shadows of the Mind: A Search for the Missing Science of Consciousness*. Oxford: Oxford University Press, 1996. Penrose's proposal for quantum mechanical effects in microtubules as a physical basis for consciousness.

Pinker, Steven. *How the Mind Works*. New York: W. W. Norton, 1997. Although the book may not fully live up to its title, it does offer a voluminous and wide-ranging set of speculations on evolution, language, and human nature.

Plato. *Phaedo*. Oxford: Oxford University Press, 2009. Circa 380 B.C. Crucial for understanding both Plato and his later influence on Medieval philosophy. Available in many editions. Downloadable at http://classics.mit.edu/Plato/phaedo.html.

———. *Republic*. New York: Dover, 2000. Circa 380 B.C. The core Platonic dialogue on social and political philosophy, building from an analogy with mind. Available in many editions. Downloadable at http://classics.mit.edu/Plato/republic.html.

Polger, Thomas W. "Zombies Explained." In Don Ross, Andrew Brook, and David Thompson, eds., *Dennett's Philosophy*, 259–286. Cambridge, MA: MIT Press, 2000. A further response to Dennett in defense of the point and force of zombie thought experiments.

Popper, Karl. "Science: Conjectures and Refutations." In Karl Popper, *Conjectures and Refutations*. New York: Harper and Row, 1968, 43–78. Reprinted in Patrick Grim, *Philosophy of Science and the Occult*, 104–111. Albany, NY: SUNY Press, 1990. Popper outlines his falsifiability criterion of demarcation, with Freud as a major target.

Purves, Dale and R. Beau Lotto. *Why We See What We Do*. Sunderland, MA: Sinauer Associates, 2003. A fascinating series of studies in psychology of perception.

Putnam, Hilary. "Meaning and Reference." *Journal of Philosophy* 70, no. 19 (1973): 699–711. Putnam's clearest presentation of his Twin Earth thought experiment.

———. “Psychological Predicates.” In W. H. Capitan and D. D. Merill, Eds. *Art, Mind, and Religion*, 37–48. Pittsburgh: University of Pittsburgh Press, 1967. One version of Putnam’s multiple instantiation argument.

Rahula, Walpola. *What the Buddha Taught*. New York: Grove Press, 1974. An easy introduction to the essentials of Buddhism.

Ramachandran, V. S. *A Brief Tour of Human Consciousness*. New York: Pi Press, 2004. A first chapter on Ramachandran’s work on phantom limbs is followed by accessible chapters that include Ramachandran’s theory of synesthesia.

Ramachandran, V. S. and Sandra Blakeslee. *Phantoms in the Brain: Probing the Mysteries of the Human Mind*. New York: Quill, 1988. Includes Ramachandran’s work on phantom limbs, but also a number of other chapters on mind and brain, both clinical and speculative.

Ratey, John J. *A User’s Guide to the Brain*. New York: Vintage, 2001. Research rich and wide-ranging but written with journalistic flair.

The Rig Veda. Ed. and tr. Wendy Doniger. New York: Penguin, 2005. The earliest of the four Hindu Vedas, circa 1200–900 BC.

Russell, Bertrand. *A History of Western Philosophy*. New York: Touchstone, 2008. First published in 1945. This is a widely condemned for its personal philosophical biases, and valuable for precisely that reason. Includes Russell’s vibrant and compelling take on Descartes, Spinoza, Leibniz and more.

Sacks, Oliver. *An Anthropologist on Mars: 7 Paradoxical Tales*. New York: Vintage, 1996. The source for the colorblind painter, mentioned briefly in the lectures. Sacks should be on every bookshelf; his points are always as illuminating as his writing is entertaining.

———. *Musicophilia: Tales of Music and the Brain*. New York: Vintage, 2007. As fascinating and accessible as all of Sacks’s volumes, particularly good on auditory hallucination, synesthesia, and musical therapy.

———. *The Man Who Mistook His Wife for a Hat*. New York: Touchstone, 1970. These lectures try to go beyond Sacks because of his familiarity, but his cases and reflections are well worth re-reading.

Searle, John. "Minds, Brains, and Programs," *The Behavioral and Brain Sciences* 3, no. 3 (1980): 417–457. Widely anthologized. Searle's classic presentation of the Chinese room thought experiment.

Shannon, Claude. "A Mathematical Theory of Communication." *The Bell System Technical Journal*, 27, nos. 3 and 4 (July, October 1948): 379–423, 623–656. Also available as *The Mathematical Theory of Communication*. Chicago: University of Illinois Press, 1971. The seminal text for all information theory.

Skinner, B. F. *Science and Human Behavior*. New York: The Free Press, 1965. The most complete source for Skinner on behaviorism.

Sperry, Roger W. "Cerebral Organization and Behavior: The split brain behaves in many respects like two separate brains, providing new research possibilities." *Science* 133, issue 3466 (1961): 1749–1757. The classic source of split-brain studies.

Spinoza, Baruch. *Ethics*. Ed. Edwin Curley. New York: Penguin, 2005. First published 1677. Bertrand Russell called Spinoza "the noblest and most lovable of the great philosophers," which also says a lot about Russell. Available in many editions.

Strawson, Galen, et al. *Consciousness and Its Place in Nature: Does Physicalism Entail Panpsychism?* Exeter, UK: Imprint Academic 2006. A collection in which Strawson outlines and defends his panpsychism against a number of critics.

Stich, Steven and Ted A. Warfield, eds. *The Blackwell Guide to Philosophy of Mind*. Malden, MA: Blackwell, 2003. A compendium of contributed pieces by contemporary philosophers on topics in philosophy of mind.

Thagard, Paul. *Hot Thought: Mechanisms and Applications of Emotional Cognition*. Cambridge, MA: MIT Press, 2006. A well-developed proposal for incorporating emotional reflection as a crucial part of practical and critical thinking.

Tononi, Guilio. "Consciousness as Integrated Information: A Provisional Manifesto." *Biological Bulletin* 215, no. 3 (2008): 216–242. Tononi's outline of consciousness as integrated information.

Tononi, Guilio. *Phi: A Voyage from the Brain to the Soul*. New York: Pantheon, 2012. A charmingly visual and literary tour through Tononi's ideas on consciousness, neurophysiology and information.

Turing, Alan M. "Computing Machinery and Intelligence." *Mind*, new series 59, no. 236 (1950): 433–460. An edited version reprinted in Douglas Hofstadter and Daniel Dennett, *The Mind's I*, New York: Basic Books, 2000, 53–66. Turing's classic piece outlining the Turing test, excerpted without some of the technical details in the Hofstadter and Dennett version.

Upanishads. Tr. Patrick Olivelle. Oxford: Oxford University Press, 2008. Sacred Hindu texts circa 800–200 B.C.

Velman, Max and Susan Schneider, eds. *The Blackwell Companion to Consciousness*. Malden, MA: Blacwell, 2007. A valuable collection of contributed pieces, often very technical.

Watson, James D. *The Double Helix*. New York: New American Library, 1968. A wonderfully personal account of the race for the structure of DNA, with telling reflections on the very human nature of scientific exploration.

Watson, John B. *Behaviorism*. New York: W. W. Norton 1970. First published 1924. In his own words.

Wegner, Daniel. *White Bears and Other Unwanted Thoughts: Suppression, Obsession, and the Psychology of Mental Control*. New York: Guilford Press, 1989. An accessible psychological guide to the difficulty of thought suppression.

Whitman, Walt. "Manly Health and Training." *Walt Whitman Quarterly Review*, 33, nos. 3–4 (2016): 184–310. Walt Whitman's newly discovered 1858 writings on exercise, in a special two-number issue of the journal.

Wise, Michael O, Martin G. Abegg, & Edward M. Cook, eds. *The Dead Sea Scrolls: A New Translation*. San Francisco: Harper, 2005. The definitive English translation of the complete scrolls, with helpful glossary and introductions.

Wittgenstein, Ludwig. *The Blue and Brown Books*. New York: Harper & Row, 1965. Dictated to his students between 1933 and 1935, circulated among his followers in blue and brown wrappers.

———. *Philosophical Investigations*. Hoboken, NJ: Wiley-Blackwell, 2009. Assembled and ordered by his students from Wittgenstein's notes shortly after his death in 1951. The definitive corpus for the later Wittgenstein.

———. *Wittgenstein's Lectures on the Foundations of Mathematics Cambridge 1939*. Ed. Cora Diamond. Chicago: University of Chicago Press, 1989. Compiled from notes by his students, effective transcripts of the Wittgenstein lectures and Wittgenstein-Turing exchanges of 1939.

Wundt, Wilhelm. *An Introduction to Psychology*. Tr. Rudolf Pinter. Blank Spots Publishing, 2014. Wundt's own introduction to his experimental approach to consciousness.

Image Credits

Page No.

Title Page .... © iyazz/iStock/Thinkstock
4...© patrice6000/Shutterstock
6................................. © life_in_a_pixel/Shutterstock
11 .... © belushi/Shutterstock
13 .... © Tomwang112/iStock/Thinkstock
14 .... © MidoSemsem/Shutterstock
16 .... © romkaziStock/Thinkstock
18 .... © Anastasios71/Shutterstock
20 .... © chatsimo/iStock/Thinkstock
24 .... © Igor Chus/Shutterstock
26 .... © designbydx/iStock/Thinkstock
28 .... © Alvaro Puig/Shutterstock
31 .... © Pushish Images/Shutterstock
33 .... © Ingram Publishing/iStock/Thinkstock
36 .... © Rawpixel.com/Shutterstock
38 .... © maxpetrov/Shutterstock
41 .... © A and N photography/Shutterstock
44 .... © Elnur/Shutterstock
46 .... © lzf/iStock/Thinkstock
48 .... © Andrew Burgess/Shutterstock
50 .... © Anastasios71/Shutterstock
51 .... © eldeiv/Shutterstock

56 .... © Brendan Howard/Shutterstock
58 .... © Georgios Kollidas/Shutterstock
61 .... © Blamb/Shutterstock
63 .... © MarcoMarchi/iStock/Thinkstock
66 .... © Lightspring/Shutterstock
68 .... © Ablestock.com/iStock/Thinkstock
74 .... © VLADGRIN/Shutterstock
76 .... © Monkey Business Images/iStock/Thinkstock
78 .... © Sebastian Kaulitzki/Shutterstock
81 .... © janulla/iStock/Thinkstock
82 .... © BananaStock/iStock/Thinkstock
86 .... © lixuyao/iStock/Thinkstock
88 .... © BananaStock/iStock/Thinkstock
91 .... © ysuel/Shutterstock
94 .... © Janie Airey/iStock/Thinkstock
96 .... © Viktor_Gladkov/iStock/Thinkstock
98 .... © SIphotography/iStock/Thinkstock
100 .... © Georgios Kollidas/Shutterstock
103 .... © okili77/Shutterstock
106 .... © Zoonar RF/iStock/Thinkstock
108 .... © tixti/iStock/Thinkstock
110 .... © Ryan McVay/iStock/Thinkstock
114 .... © Argument/iStock/Thinkstock
116 .... © OSTILL/iStock/Thinkstock
118 .... © KatarzynaBialasiewicz/iStock/Thinkstock
120 .... © Andrey_Popov/Shutterstock
126 .... © moodboard/iStock/Thinkstock
128 .... © RyanKing999/iStock/Thinkstock
130 .... ©selensergen/iStock/Thinkstock
132 .... ©Telia/Shutterstock
136 .... © ESB Professional/Shutterstock
138 .... © Lightspring/Shutterstock
140 .... © PopTika/Shutterstock
142 .... © 1000 Words/Shutterstock
148 .... © Ociacia/Shutterstock
150 .... © Willyam Bradberry/Shutterstock
156 .... © PHOTOCREO Michal Bednarek/Shutterstock

158................................© Christian Lagerek/Shutterstock
160...................................... © sarah5/iStock/Thinkstock
162...........................© andrewsafonov/iStock/Thinkstock
166...................© Wavebreakmedia Ltd/iStock/Thinkstock
170........................... © cosmin4000/iStock/Thinkstock
172......................................© iLexx/iStock/Thinkstock
178..............................© DigtialStorm/iStock/Thinkstock
180................................ © Catalin205/iStock/Thinkstock
182............................© Digital Vision/iStock/Thinkstock
188.................................. © goa_novi/iStock/Thinkstock
190.................................... © Eraxion/iStock/Thinkstock
192......................© everythingpossible/iStock/Thinkstock
200 ....© Jochen Sand/iStock/Thinkstock
202................................© BlindTurtle/iStock/Thinkstock
206 ....© shironosov/iStock/Thinkstock
210...............................© Ryan McVay/iStock/Thinkstock
212...................© Wavebreakmedia Ltd/iStock/Thinkstock
217.......................... © eugenesergeev/iStock/Thinkstock
220...................© Wavebreakmedia Ltd/iStock/Thinkstock
222..................................© Antonio Guillem/Shutterstock
225.....................................© Claudio Divizia/Shutterstock
230................................ © Ingram Publishing/Thinkstock
232.......................................© monsitj/iStock/Thinkstock
234 ....© monkeybusinessimages/iStock/Thinkstock
240.................................... © STILLFX/iStock/Thinkstock
242...............................© DimaBerkut/iStock/Thinkstock
245................................ © muratsenel/iStock/Thinkstock